Analysis of cbbL*,* nifH*, and* pufLM *in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of autotrophic and phototrophic bacteria.* Guillaume Tahon, Bjorn Tytgat, Pieter Stragier, Anne Willems. Microbial Ecology (2016) 71:131-149.

DOI: 10.1007/s00248-015-0704-6

Analysis of *cbbL*, *nifH* and *pufLM* in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of auto- and phototrophic bacteria

Guillaume Tahon, Bjorn Tytgat, Pieter Stragier, Anne Willems*

Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium

***Corresponding author:**

Anne Willems

Laboratory of Microbiology, Department of Biochemistry and Microbiology

Ghent University

Ghent, Belgium

Tel.: +32 (0) 9 264 5103

Fax: +32 (0) 9 264 5092

Email: Anne.Willems@UGent.be

To refer to or to cite this work, please use the citation to the published version:

Tahon G., B. Tytgat, P. Stragier, A. Willems. 2016. Analysis of *cbbL*, *nifH*, and *pufLM* in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of autotrophic and phototrophic bacteria. Microb. Ecol. 71:131-149.

http://dx.doi.org/10.1007/s00248-015-0704-6

Keywords

Antarctica; RuBisCO; *nifH*; *pufLM*; Bacteria; Clone Libraries

Abstract

Cyanobacteria are generally thought to be responsible for primary production and nitrogen fixation in the microbial communities that dominate Antarctic ecosystems. Recent studies of bacterial communities in terrestrial Antarctica, however, have shown that Cyanobacteria are sometimes only scarcely present, suggesting other bacteria presumably take over their role as primary producers and diazotrophs. The diversity of key genes in these processes was studied in surface samples from the Sør Rondane Mountains, Dronning Maud Land, using clone libraries of the large subunit of ribulose-1,5-biphosphate carboxylase/oxygenase (RuBisCO) genes (*cbbL, cbbM*) and dinitrogenase-reductase (*nifH*) genes. We recovered a large diversity of non-cyanobacterial *cbbL* type IC in addition to cyanobacterial type IB, suggesting that non-cyanobacterial autotrophs may contribute to primary production. The *nifH* diversity recovered, was predominantly related to Cyanobacteria, particularly members of the *Nostocales*. We also investigated the occurrence of proteorhodopsin and anoxygenic phototrophy as mechanisms for non-Cyanobacteria to exploit solar energy. While proteorhodopsin genes were not detected, a large diversity of genes coding for the light and medium subunits of the type 2 phototrophic reaction center (*pufLM*) was observed, suggesting for the first time, that the aerobic photoheterotrophic lifestyle may be important in oligotrophic high altitude ice-free terrestrial Antarctic habitats.

Introduction

The Antarctic continent is dominated by extreme environmental conditions, including limited organic nutrients, low humidity, frequent freeze-thaw and wet-dry cycles, rapid drainage, low and transient precipitation and low thermal capacity of the substratum [1]. Ice-free regions are limited to a minute proportion (~0.32%) of the continent's surface area [2]. The Sør Rondane Mountains, located in Dronning Maud Land, East-Antarctica and home of the Belgian Princess Elisabeth Station (PES) contain an important number of ice-free mountain tops and nunataks [3].

Antarctica has a very limited fauna and flora, with just two native species of flowering plants, occurring only on the Peninsula [4, 5]. Surface substrates near the PES consist mostly of weathered rocks with relatively little organic material, given the absence of vascular plants. It is assumed that Cyanobacteria, with their ability to fix carbon dioxide and nitrogen, are the main primary producers in this barren region [6, 7]. Recent studies however, have shown that Cyanobacteria are not always highly abundant in vegetated and fell-field sites on the Peninsula and in mineral soils of the McMurdo Dry Valleys [8-10] and in such cases, other bacteria can be assumed to take over their role in carbon and/or nitrogen fixation.

Of several mechanisms for carbon dioxide fixation, the Calvin-Benson-Bassham (CBB) cycle is considered to be the most important autotrophic pathway [11]. The enzyme responsible for the actual CO_2 fixation in the CBB cycle is ribulose-1,5-biphosphate carboxylase/oxygenase [12]. Four different natural types of RuBisCO are known: Type I (*cbbL* gene) is subdivided in to a green and a red subgroup, each containing two subclasses [11]. In the green subgroup, *cbbL* IA is predominantly found in Alpha-, Beta- and Gammaproteobacteria, while type IB is found in Cyanobacteria, green algae and green plants [13]. The red group contains type IC, found in Alpha-, Beta- and Gammaproteobacteria [13, 14], but recently also in Verrucomicrobia [15], Firmicutes [16] and Actinobacteria [17, 18], and type ID, found in non-green algae [13, 14]. Some authors have suggested that the type I found in Verrucomicrobia and Actinobacteria may form a separate subtype in the red subgroup, type IE *cbbL* [15, 17, 18]. Type II (*cbbM* gene) is found in purple non-sulfur bacteria, aerobic and facultative anaerobic chemoautotrophic bacteria and dinoflagellates [19, 20]. RuBisCO

type III has only been found in Archaea [21]. Finally, type IV RuBisCO is designated as RuBisCO-like [14] and is considered not to be involved in the Calvin cycle [22].

Besides fixation of carbon dioxide, biological nitrogen fixation, a multistep process catalyzed by the nitrogenase enzyme complex [23], is also very important in Antarctic soils [24, 25]. The nitrogenase complex is composed of two metalloproteins: dinitrogenase and dinitrogenase-reductase. The latter, encoded by *nifH*, is very useful in phylogenetic studies because of its conserved nature [26] and high level of congruence with 16S rRNA gene phylogeny [27].

Both nitrogen and carbon dioxide fixation require a substantial amount of ATP. In the poor soils of the ice-free zones in continental Antarctica, sunlight – an abundant energy source during the Antarctic summer – may therefore represent an important resource, also for non-Cyanobacteria. Over the past decades, several alternative light-harvesting mechanisms, including microbial rhodopsins and aerobic anoxygenic phototrophy, have been discovered. By far the most abundant family of microbial rhodopsins, proteorhodpsins (PR) is found in Bacteria where these membrane embedded light-driven proton pumps contribute to proton motive force [28, 29]. They have been reported from diverse marine and non-marine habitats [30]. To our knowledge, their presence in Antarctica, however, has not been reported.

The aerobic anoxygenic phototrophic bacteria (AAP) are photoheterotrophs that can use bacteriochlorophyll *a* to harvest light energy. Light-mediated CO_2 incorporation also occurs, although this activity is rather low [31] and remains controversial [32]. Diversity of AAP has been examined by screening the *pufL* and *pufM* genes, which encode the light and medium subunit of the type 2 photoreaction center [33]. Until now, AAP have been predominantly documented in the Alpha-, Beta- and Gammaproteobacteria (purple bacteria) [34, 35], although the light-harvesting complex is also present in anaerobic green non-sulfur bacteria, heliobacteria and other Firmicutes [36, 37]. Previous studies have revealed presence of AAP in various aquatic habitats, ranging from the world's oceans [38-42] to inland lakes [34, 35, 43, 44] and even Antarctic sea ice [45]. Until today, little is known about the presence and diversity of AAP in terrestrial environments, including continental Antarctica.

We hypothesized that in ice-free areas of terrestrial Antarctica, in addition to Cyanobacteria, other prokaryotic primary producers, and bacteria that exploit solar energy, may contribute to carbon and/or

nitrogen fixation. Our area of study was the Sør Rondane Mountains, in the vicinity of the PES, where a previous pyrosequencing study indicated scarcity of Cyanobacteria in some of the samples of terrestrial microbial mats [46]. The aim of this study was to investigate terrestrial samples for the presence and diversity of primary producers and other bacteria exploiting sunlight, by targeting *cbbL* and *cbbM* genes (RuBisCO), *nifH* (nitrogen fixation), *pufLM* (phototrophy with type 2 reaction centers, including AAP) and proteorhodopsin genes. This work represents the first study of these mechanisms in oligotrophic high altitude soils of the Sør Rondane Mountains and is the first report of the possible presence of AAP in terrestrial Antarctic samples.

Methods

Sample collection and environmental data

Surface samples consisting of weathered granite parent material (depth of 0 to 5 cm) were taken aseptically during the Antarctic summer of 2009 near the Princess Elisabeth Station (71°57'S, 23°20'E) at Utsteinen, Sør Rondane Mountains, East Antarctica. Upon collection, samples were frozen at -20°C until processing. Four samples were used in this study: KP15, KP43 and KP53 were collected on the Utsteinen ridge in the immediate vicinity of the PES, KP2 was collected on the north face of the nunatak 1.3 km south of the research station (Fig. 1). Sample description, coordinates, altitude, geochemical parameters (pH, water content, total organic carbon (TOC) and conductivity) and relative abundance of Cyanobacteria according to Tytgat et al. (unpublished data) are listed in Table 1.

Isolation of DNA

Seven DNA extraction methods were tested and their performance to yield maximum diversity of bacterial DNA from Antarctic terrestrial surface samples was verified by denaturing gradient gel electrophoresis of partial 16S rRNA and partial *pufM* genes. The PowerLyzer® PowerSoil® DNA isolation kit (MoBio Laboratories) performed best (Fig. S1) and was therefore used for analyzing the samples. Total genomic DNA was extracted in triplicate from 400 mg of homogenized sample, following the manufacturer's protocol. The lysis step was performed using an alternative protocol

separately provided by the manufacturer. Briefly, 500 µl phenol:chloroform:isoamylalcohol (25:24:1), 500 µl Bead Solution and 60 µl of C2 solution were added to a PowerLyzer™ Glass Bead Tube together with 400 mg of sample. Tubes were bead beaten for 10 minutes (30.0 Hz) and the supernatant was transferred to a clean 2 ml collection tube. A volume of 150 µl of component C3 was added and tubes were cooled for 5 minutes at 4°C, after which the original protocol was continued. Extracted DNA was quantified with a Qubit® 2.0 fluorometer (Life Technologies) and stored at -20°C until processing.

PCR

For every DNA extract, a PCR was performed in triplicate in a total volume of 25 µl containing 0.2 mM of each deoxynucleotide triphosphate (dNTP), 1X Qiagen PCR buffer (Qiagen), 0.625 units of Qiagen *Taq* DNA polymerase (Qiagen), 100 µM bovine serum albumin (BSA), 3 µl of template solution and a forward and reverse primer (Table 2). Amplification was performed using a Veriti thermal cycler (Life Technologies). The temperature profiles of all PCRs are shown in Table 2. For each sample and primer set, all nine PCR products (3 replicate DNA extracts x 3 replicate PCRs), showing bands of the expected size, were pooled and purified using a Nucleofast 96 PCR clean up membrane system (Macherey-Nagel) and Tecan Genesis Workstation 200 (Tecan).

Clone library construction and sequencing

Purified PCR products were cloned with a pGEM®-T Vector System II (Promega) following the manufacturer's instructions, in duplicate or triplicate. Competent *Escherichia coli* JM109 cells were transformed with the ligation product and screened using blue/white coloration. For each sample, 150 white transformants of each PCR type were purified by streaking. To release plasmid DNA, cells were suspended in 15 µl of MilliQ water and lysed by heating to 100°C (10 min). Inserts were amplified using the T7/SP6 primer set (Promega) in a 25 µl reaction mixture containing 1 µl of DNA solution, 1X Qiagen PCR buffer (Qiagen), 0.2 mM of each dNTP, 0.1 µM of each primer, 0.3125 units of Qiagen *Taq* polymerase (Qiagen) and 0.5 µl $MgCl_2$ (25 mM). Amplification was carried out as follows:

95°C (5 min), 3 cycles of 95°C (1 min), 49°C (2 min 15 sec), 72°C (2 min 15 sec) and 30 cycles of 95°C (35 sec), 49°C (1 min 15 sec), 72°C (1 min 15 sec). A final extension at 72°C (7 min) and subsequent cooling at 4°C completed the reaction. PCR products of the expected size were purified as described above and gene fragments were sequenced with primers T7 and SP6 and the original amplification primers (Table 2) using a BigDye Xterminator™ purification kit (Applied Biosystems) and an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).

Analysis of clones

For each functional gene, a custom-made database was prepared in BioNumerics 7.5 (Applied Maths) by downloading all related sequence records from the NCBI and IMG (https://img.jgi.doe.gov/) [47] databases as available per December 1st 2014. Redundant sequences were not removed not to lose metadata information on the habitats reference sequences originated from. Records were manually checked to eliminate low-quality sequences (i.e. presence of stop codons, ambiguous bases, indels). Newly obtained sequences were added to the corresponding BioNumerics database using the Assembler module. For all clones, between 4 and 8 overlapping sequences were assembled. Vector sequences and amplification primer sequences were trimmed off and sequences were manually curated. For additional quality curation, nucleotide sequences were translated into amino acids using MEGA 6 [48]. Putative chimeric sequences (detected using the Uchime module in Mothur [49, 50]), sequences with no similarity to our genes and sequences that were too short (≥ 1 AA) or contained stop codons were removed from the dataset before further analysis. For rarefaction analysis, binning at both 95% and 100% nucleotide sequence identity and 100% protein sequence identity was performed using the "Fill field with cluster number" option in BioNumerics 7.5. For phylogenetic analyses of *cbbL*, *nifH* and *pufLM*, all unique protein sequences were included. They are referred to as operational RuBisCO units (ORU), operational nitrogenase units (ONU) and operational puf units (OPU). For each gene, a first amino acid alignment was made with all sequences present in our database (*nifH*: 46.371; *pufLM*: 3.706; *cbbL* type IA, IB, IC: 8.004), using Clustal Omega [51, 52]. Alignments were trimmed to the size of our cloned fragments and visually inspected, excluding all non-overlapping sequences or sequences that were too short (≥ 1 AA) from further analysis. These alignments were used to construct

184 a maximum likelihood (ML) tree (1.000 bootstrap replicates) with FastTree software [53] using the
185 Whelan and Goldman evolutionary model and the discrete gamma model with 20 rate categories.
186 Analysis of the *cbbL* IA/IB sequences, together with reference data, resulted in a final tree containing
187 2.103 protein sequences. For *cbbL* IA/IC, *nifH* and *pufLM*, these trees contained 3.610, 26.370 and
188 3.529 protein sequences respectively (data not shown). From the resulting phylogenetic trees, closest
189 relatives of our newly obtained clone sequences as well as representative sequences from the entire
190 tree were selected in order to prepare a smaller tree representing the initial complete tree, following the
191 same protocol. Trees were visualized using the iTOL software [54, 55]. Clone sequences sharing the
192 same nearest neighbor were grouped into clusters.
193
194 **Statistical analyses**
195 Statistical analyses of clone sequences were performed using the Vegan package [56] in R
196 (http://cranr-project.org). The total expected number of OTUs was determined by rarefaction analysis.
197 For all clone libraries, different indices, based on derived protein sequences, were calculated,
198 including evenness (Pielou), species richness (Chao1), and Bray-Curtis dissimilarity.
199
200 **Accession numbers**
201 The sequences determined in this study have been deposited in the National Center for Biotechnology
202 GenBank database under the accession numbers KT154121 to KT154253 (IA/IB *cbbL* sequences),
203 KT154254 to KT154456 (IA/IC *cbbL* sequences), KT154018 to KT154120 (*nifH* sequences) and
204 KT154457 to KT154682 (*pufLM* sequences).
205
206 **Results**
207 Four terrestrial samples (Table 1), collected in the proximity of the Belgian Princess Elisabeth Station,
208 were examined. To target different subtypes of the *cbbL* gene, coding for RuBisCO type I, two primer
209 sets were used, RubIgF/R and K2f/V2r which robustly capture subtypes IA/IB and IA/IC, respectively
210 (Fig. S2, Table 2) [21, 57-62]. A total of 353 clones of subtypes IA/IB and 444 of types IA/IC were

obtained (Table 3). The *cbbM* gene (type II RuBisCO) failed to amplify from all four samples even though a recently designed primer set reported to target a broad diversity of the gene was used (Table 2) [63]. For *nifH* and *pufLM*, a total of 339 and 317 clones were obtained in this study (Table 3). Proteorhodopsin genes failed to amplify from all four samples.

When applying OTU binning at 95% DNA similarity, for all clone libraries rarefaction curves reached saturation (Fig. 2), reflecting that at a 5% DNA divergence level the number of OTUs is relatively limited. Indeed, for *cbbL* subtypes IA/IB, *cbbL* subtypes IA/IC, *nifH* and *pufM* at 95% DNA sequence similarity 14, 43, 3 and 21 OTUs were recovered (Table 3). However, analysis of unique protein sequences or unique DNA sequences indicated that not all diversity was sampled as graphs were still rising (Fig. 2), in line with overall coverage values of about 35 to 50% (Table 3).

For the *cbbL* gene the two primer sets used to amplify subtypes IA/IB and IA/IC target slightly different but overlapping regions of the gene that were therefore investigated separately (Fig. S2, Table 2). For both primer sets, at least 98 high quality sequences were retained per sample, except for samples KP15 and KP53, where fewer sequences were obtained for one of the primer sets (Table 3). The coverage based on estimated species richness (Chao1) for the separate samples was only between 23.1% and 79.2% for IA/IB and between 27.1% and 74.7% for IA/IC, whereas the total coverage of the IA/IB and IA/IC libraries was only 35.4% and 40.7%. The overall evenness of the clone libraries was quite high in all the cases. Only for *cbbL* clone libraries of sample KP43, the value was lower. Finally, Bray-Curtis dissimilarity analysis revealed that, with the available data, very little overlap exists between the data from all terrestrial samples (Table 3, Fig. 3).

For *cbbL* IA/IB, the 353 clones were binned into 98 ORUs that had a unique amino acid sequence. Similarly, for types IA/IC, a total of 158 ORUs were defined from 444 clones (Table S1). For both subtypes, most ORUs comprised only one or two derived protein sequences and little overlap existed between ORUs originating from different terrestrial samples (Fig. 3). After maximum likelihood analysis of the *cbbL* IA/IB sequences, together with reference data, all 98 ORUs grouped into 9 visual clusters and 3 separate ORUs dispersed between other available sequence data (Fig. 4). Clusters 4 to 9 and two separate ORUs grouped with cyanobacterial type IB. The number of type IB sequences recovered from samples KP2, KP15, KP43 and KP53, was 115, 35, 15 and 4, respectively (Table 4).

Blast searches revealed none of them to be similar to eukaryotic *cbbL*. They made up 50 ORUs that showed very little overlap between samples. In fact, only two visual clusters (clusters 6 and 7) (Table 4) in the ML tree (Fig. 4) contained ORUs recovered from multiple samples. Most type IB ORUs contained sequences originating from a single sample. The type IB sequences showed a clear taxonomic grouping of Cyanobacteria with clusters of *Oscillatoriales*, *Nostocales* and *Chroococcales* sequences (Fig. 4) and two clone clusters (4 and 8) showed a close affiliation to *Chroococcales cbbL* IB sequences. Separate ORU KP2.RuBisCO.AB.Clone89 and clusters 5 – 7, all from samples KP43 and KP53, grouped with sequences of *Nostocales*, while cluster 9 and singleton KP15.RuBisCO.AB.Clone61 grouped with the *Oscillatoriales*. Our type IB sequences mostly appear cosmopolitan as their closest relatives originate from samples taken in a variety of ecosystems all over the world. A few (e.g. KP15.RuBisCO.AB.Clone61, Cluster 8, Cluster 9), however, have a sequence from Antarctica or other cold environments as closest neighbor (Table 4) [21, 64].

Cluster 3 plus a separate ORU grouped with type IA and these sequences showed a close affiliation with the *cbbL* type IA sequence from *Bradyrhizobium* sp. BTAi1 (Fig. 4). Remarkably, despite the amplification with primers targeting only type IA/IB, two clusters (1 and 2) grouped with type IC sequences (Fig. 4), in particular with unnamed clone sequences obtained from Chinese arid (Accession no. FR849177, FR849160, FR849127) and German agricultural (Accession no. AY572135) soil.

A similar analysis of the clone sequences for *cbbL* IA/IC grouped all 148 ORUs into 16 clusters and 26 separate ORUs (Fig. 5, Table 4). Only 1 separate ORU, containing one sequence (KP15.RuBisCO.AC.Clone113), grouped with sequences of type IA. It grouped with the *cbbL* IA sequence of *Bradyrhizobium* sp. BTAi1, as was the case for all type IA sequences obtained with the other primer set. The remaining 147 ORUs, containing 443 sequences, all grouped with *cbbL* IC (Table 4). Apart from cluster 12 which grouped with *Nocardia* species and cluster 16, which grouped with *Nitrosospira* species, most new cloned sequences grouped with sequences of other uncultured bacteria (Fig. 5, Table 4).

For amplification of a ~360 bp fragment of the *nifH* gene, a degenerate primer set, previously described by Poly and colleagues [65], was used. The primers target two of the three main conserved *nifH* regions, that have been used to design nearly all existing *nifH* primers [66]. For sample KP53, no

successful amplification could be obtained. For the other samples, amplification was successful: 115, 119 and 105 clones were retained for samples KP2, KP15 and KP43 respectively (Table 3). Overall, 62 ONUs were obtained. Bray-Curtis dissimilarity again revealed hardly any overlap between the data from different samples, except for samples KP2 and KP15 (Table 3), which shared 4 protein sequences (Fig. 3), of which one (ONU 1) was recovered from all three samples (KP2: 85 clones, KP15: 92 clones, KP43: 5 clones) (Table S1). The overall coverage of the *nifH* clone libraries, based on the Chao1 parameter, was highest for sample KP43 (45.8%), but dropped to 44.4% and 27.5% for clone libraries of KP15 and KP2 respectively. Evenness was low for all samples (Table 3). After maximum likelihood analysis of the protein sequences, together with our *nifH* database sequences, revealed all our ONUs to group with Cyanobacteria (Fig. 6). They formed one cluster together with several *Nostocales* reference strains, especially *Nostoc* (Table 4).

For amplification of *pufLM* genes a degenerate primer set was used (Table 2) that amplifies a ~1500 bp long fragment, providing additional sequence information compared to other primer sets that only amplify small parts of the *pufL* or *pufM* gene. All samples, except for KP2, showed successful amplification. Cloning resulted in 98, 98 and 121 positive sequences for samples KP15, KP43 and KP53 respectively (Table 3).

Analysis of the complete ~1500 bp fragment revealed 193 OPUs with a unique amino acid sequence (data not shown). To date, most available sequences comprise a smaller part of the *pufM* gene. Therefore, to broaden phylogenetic analysis by comparing our new sequences with as much reference data as possible, only this small part of the *pufM* gene, covering 58 AA, was used for further analysis. This increased the reference data in the phylogenetic analysis by 17 times, while reducing our sequences to 49 unique OPUs (Table S1), 46 of which consisted of sequences originating from just one of the terrestrial samples. Of the other 3 OPUs, two (OPU 22 and OPU 40) were shared between three terrestrial samples and one (OPU 2) between samples KP15 and KP53 (Fig. 3).

Bray-Curtis analysis indicated samples KP43 and KP53 to show only 57% dissimilarity, while all other relationships show a very high dissimilarity (Table 3). Evenness was highest for the KP15 library (81.4%) and 49.6% and 52.5% for the KP43 and KP53 libraries respectively. Coverage of the

libraries was 62.8%, 91.4% and only 30.4% for samples KP15, KP43 and KP53, respectively (Table 3).

For phylogenetic analysis, a maximum likelihood tree (1.000 bootstraps), was generated (Fig. 7), in which the 49 OPUs could be grouped in 6 visual clusters and five separate OPUs (Table 4). Sequences obtained from sample KP15 proved to be the most diverse, with OPUs belonging to all six clusters and all five separate OPUs. Sequences from samples KP43 and KP53 were much less diverse, belonging to only two and three of the visual clusters, respectively (Table 4).

Only for cluster 5 and singleton KP15.pufLM.Clone55, the closest affiliated unknown *pufM* sequences (Accession no. KC900142 and KC900120) also originates from Antarctica (Fig. 7), however, they were not found in a terrestrial environment, but in surface water from Great Wall Cove, King George Island, Antarctica. Furthermore, cluster 1 and singleton KP15.pufLM.Clone13 (Fig. 7) group with unknown *pufM* sequences originating from Chinese paddy soil samples [67-69]. Clusters 2 and 3, and singletons KP15.pufLM.Clone14, -Clone55 and –Clone78 grouped into a larger cluster (from *Sphingomonas* sp. PB56 [AY853583] to *Sphingomonas echinoides* ATCC 14820 [NZ_JH584235]) (Fig. 7) of *pufM* sequences belonging mainly to organisms of the alphaproteobacterial orders of the *Rhizobiales*, originating from terrestrial and aquatic environments, and *Sphingomonadales*, predominantly originating from aquatic environments.

Discussion

Based on previous research, where Cyanobacteria were reported in small numbers in some exposed Antarctic communities [8-10], we hypothesized that other members of the bacterial community may take their role as primary producers in the oligotrophic soils of the Antarctic continent. We investigated this in four samples from soils near the Princess Elisabeth Station in the Sør Rondane Mountains. As part of a larger biogeographic study these samples were recently included in Illumina sequencing of 16S rRNA genes (Tytgat et al, unpublished data) and preliminary analyses indicate that Cyanobacteria make up 0.04 to 51.85% of the reads recovered from our four samples (Table 1). We studied the diversity of genes encoding RuBisCO as a marker for phototrophic and chemolithotrophic primary producers and the nitrogenase gene *nifH* as a marker for nitrogen fixation. The scarcity of

organic matter and abundant availability of sunlight would favor bacteria that are able to exploit the latter resource. Therefore we also investigated the diversity of type 2 phototrophic reaction centers (*pufLM* genes), which are also found in aerobic photohetertrophs, and PR light harvesting systems.

Few studies have reported the presence and diversity of *cbbL* and/or *cbbM* genes in Antarctica, most of them investigating aquatic ecosystems [21, 70, 71], leaving the ice-free terrestrial regions unexplored. Here, two different degenerate primer sets were used to target the diversity of RuBisCO type I (*cbbL* gene) in soils near the Princess Elisabeth Station. The first amplified both types IA and IB, the second types IA and IC. This division of RuBisCO type I into different subtypes can be seen clearly in the phylogenetic trees (Fig. 4 and Fig. 5). The *cbbL* t*ype* IC, previously found in Alpha-, Beta- and Gammaproteobacteria, Verrucomicrobia, Firmicutes and Actinobacteria, was the most frequently recovered type in all samples (202 ORUs) (Table 4) and showed very high diversity, making up 18 clusters and 26 separate ORUs (Fig. 4 and Fig. 5, Table 4), suggesting the presence, in the terrestrial areas near the Belgian base, of multiple non-cyanobacterial autotrophs, presumably belonging to a wide range of taxa. Remarkably, the primer set targeting types IA/IB also picked up a large number of type IC sequences from samples KP43 and KP53 (108 and 72 sequences respectively) (Table 4), indicating that the primer set, although successfully amplifying only *cbbL* IA/IB in previous studies [58, 59, 72-74], is not 100% specific for these two subtypes.

Type IA *cbbL*, typical of Alpha-, Beta- and Gammaproteobacteria, was amplified only in sample KP15 with a very limited diversity of only 4 ORUs (5 DNA sequences), despite the use of two primer sets that target this type (Table 4). Possible reasons for this low diversity of type IA may be a primer bias, the absence or low abundance of the targeted *cbbL* type IA. Based on literature data, the latter may be the case, as both primer sets were shown to display only limited bias amongst the different *cbbL* types [62] and they successfully amplified type IA in other studies [21, 57, 58, 60-62]. A further possibility is the presence of slightly different type IA genes that were not picked up by the primers. Primer mismatch is a weakness for all PCR-based diversity surveys and may cause a portion of the actual diversity to be missed.

348 Finally, a large diversity of cyanobacterial *cbbL* type IB was recovered (Fig. 4). While there was very
349 little overlap between samples, KP2 and KP15 yielded the most diversity, in line with the higher
350 relative abundance of Cyanobacteria in those samples (Tables 1 and 4).
351 Taken together, the diversity of RuBisCO type I sequences recovered with two primer sets from the
352 terrestrial Antarctic samples revealed a larger relative diversity of non-cyanobacterial primary
353 producers (202 ORU of type IC and 4 ORU of type IA) than Cyanobacteria (50 ORU), confirming our
354 hypothesis that these groups may potentially contribute to primary production in these systems.
355 Furthermore, the presence of *cbbL* type IA sequences similar to those retrieved from the
356 alphaproteobacterium *Bradyrhizobium* sp. BTAi1, type IB sequences from *Oscillatoriales*, *Nostocales*
357 and *Chroococcales*, and type IC from *Nitrosospira* and *Rubrivivax* corroborates previous reports from
358 Antarctic or subglacial environments [21, 75].
359 It should be noted that while many of the clone sequences show relatively high similarity to sequences
360 of cultivated bacteria, this does not imply taxonomic relatedness. The *cbbL* phylogeny shows a high
361 degree of incongruency with the 16S rRNA phylogeny, due to horizontal gene transfer and gene
362 duplication associated with differential gene loss [76]. Considering the type of habitat related
363 reference sequences originate from, no clear grouping of cold habitat sequences was apparent for any
364 of the *cbbL* types (Table 4).
365 Previous studies reported the presence of type II RuBisCO in Antarctica [70, 71], however, despite the
366 use of a primer set recently shown to amplify a 700-800 bp fragment of the *cbbM* gene in a broad
367 variety of bacteria [63], no positive amplification or no amplicon of the correct size could be obtained
368 from our samples (data not shown). These results imply that type II RuBisCO sequences are absent or
369 below the detection limit in our samples or that they do not match the primers used.
370 We also studied the diversity of dinitrogenase-reductase (*nifH*) genes to assess to what extent nitrogen
371 fixation might be performed by non-cyanobacterial taxa. The primer set used targets a broad variety of
372 Alpha-, Beta- and Gammaproteobacteria, Gram positive bacteria and only a small percentage of
373 known cyanobacterial *nifH* sequences [65, 66]. All obtained sequences, however, showed a close
374 relationship to known *nifH* sequences belonging to *Nostocales* cyanobacteria and especially the
375 heterocystous cyanobacterium *Nostoc*. No non-cyanobacterial sequences were recovered, suggesting

that non-Cyanobacteria contribute little to nitrogen fixation in these terrestrial Antarctic samples. Although we recovered mostly *Nostocales*-affiliated *nifH* sequences, this does not imply these are the main nitrogen fixers because the primers used did not target the majority of Cyanobacteria [66]. Since rarefaction curves based on unique sequences did not reach saturation (Fig. 2) and coverage estimates are below 50% (Table 3), it is possible that other Cyanobacteria as well as non-Cyanobacteria may contribute to nitrogen fixation in the terrestrial Antarctic. The presence of *Nostocales* Cyanobacteria, and in particular *Nostoc*, in our terrestrial Antarctic samples as suggested by *nifH* findings, is corroborated by the *cbbL* type IB data (Fig. 4). *Nostoc* is indeed frequently found in Antarctic environments [7, 9, 77, 78] as well as other terrestrial ecosystems characterized by aridity and nutrient limitations [79]. Its success in these extreme environments is due to its ability to remain desiccated for extended periods of time and, after rehydration, recover its metabolic activity completely within hours to days [79]. Moreover, *Nostoc* can withstand repeated cycles of freezing and thawing, a typical Antarctic environmental condition [80, 81]. No particular psychrophilic nitrogenase associations could be detected: the most closely related *nifH* sequences originated from a variety of ecosystems from all over our planet (Table 4).

In the present study, two mechanisms to harvest sunlight were examined. The first, proteorhodopsin, has been almost exclusively studied in aquatic ecosystems, including from Arctic and Antarctic marine environments [82-84]. Nevertheless, PR-like variants have been found in Siberian permafrost samples [30], suggesting that they could also play a role in continental Antarctica. However, proteorhodopsin genes were not detected in any of the samples tested here, despite the use of multiple primer sets (Table 2), amplifying proteorhodopsin in a wide range of bacteria [82, 83, 85, 86]. Possibly proteorhodopsins are not abundant or maybe not present at all in these soils. Another likely explanation may be that the terrestrial Antarctic variants of proteorhodopsin are not targeted by these primer sets: as most reference data are from marine systems, the available primers may not fully capture terrestrial proteorhodopsin diversity.

A second mechanism to harvest sunlight that was investigated here is anoxygenic phototrophy using type 2 reaction centers, encoded by *pufLM* genes and found in Chlorofexaceae, Purple sulfur bacteria, Purple non-sulfur bacteria and AAP [87]. Diverse *pufLM* genes were recovered from all but one

samples, indicating that anoxygenic phototrophs are present. We did not recover sequences representative of strictly anaerobic photoautotrophic Purple sulfur bacteria or Chloroflexi and only cluster 6 grouped among sequences of Purple non-sulfur bacteria (Fig. 7). Most of our clusters grouped among diverse lineages originating from aerobic photoheterotrophic taxa of Alpha- and Betaproteobacteria referred to as AAP (Fig. 7). Studies on AAPs in soils are extremely rare to date. Only Feng and colleagues [67-69, 88, 89] investigated their presence in paddy soils from China and in Arctic soils. Other studies to date focused on aquatic ecosystems [41, 45, 90], where it has been shown that AAP represent an important part of the total bacterial community [91]. This work, with the Sør Rondane Mountains as area of study, represents the first report of AAP bacteria in Antarctic soils.

Since to date most available sequences only comprise a small part of the 3' side of the *pufM* gene, this 58 AA long region was used for analysis. When expanding sequence data to the larger, originally obtained *pufLM* protein sequences, these affiliations with known AAP bacteria can still be seen, indicating that this 58 AA long region of *pufM* is indeed a useful indicator for *pufLM* phylogeny. In the terrestrial samples, a high diversity of *pufM* genes was present and clear differences were observed between the samples. Combined with the samples' conductivity data (Table 1) – and consequently salinity – a tentative correlation with *pufM* diversity could be seen, as diversity decreased when conductivity increased. This inverse correlation between salinity and *pufM* diversity was consistent with previous observations in aquatic ecosystems [44, 90]. The opposite correlation, however, has also been found [34].

Most of the *pufM* diversity recovered here affiliated with alphaproteobacterial *pufM* (Table 4, Fig. 7), an observation made previously in the Arctic [88] and Antarctic [43, 45], but also in Chinese paddy soils [68]. Furthermore, some of the alphaproteobacterial-like OPUs were found grouping with *pufM* sequences originating from *Sphingomonadales* bacteria, an order that has previously been shown to be present in soils of the Sør Rondane Mountains [92]. Gammaproteobacterial-like *pufM* sequences were also found in our samples (Table 4, Fig. 7), but contributed less to the general *pufM* diversity, as was the case for Arctic soils [88], but not Antarctic sea ice and seawater, where no *pufM* of the gammaproteobacterial group was detected [45]. Again, no clear psychrophilic *pufM* groupings could be deduced from the geographic origin of these closest relatives, as they originate from marine and

terrestrial environments from all over the world, a trend that could also be seen for the other functional genes (Table 4).

Conclusion

Overall, the data presented in this study suggest that, in soils in the vicinity of the Princess Elisabeth Station, a broad diversity of microorganisms harboring RuBisCO genes is present. Type IA was only scarcely recovered and all sequences showed close affiliation to the *cbbL* type IA sequence of *Bradyrhizobium* sp. BTAi1. Non-cyanobacterial type IC was dominantly retrieved and appeared to be more diverse than cyanobacterial type IB in the bacterial autotrophic communities near the research station, and may potentially contribute to the input of organic matter into the oligotrophic Antarctic systems. The *nifH* diversity recovered was low, as only sequences affiliated with *Nostocales* were recovered from our samples, suggesting non-Cyanobacteria are not important contributors to nitrogen fixation in the poor Antarctic soils. Our study of mechanisms to harvest solar energy could not detect rhodopsin genes. However, it did, for the first time, show the presence of aerobic anoxygenic phototrophic bacteria in Antarctic soils, suggesting that photoheterotrophy may be a useful life strategy during the austral summer.

The functional genes studied here did not appear to show Antarctic types: the majority of sequences grouped among neighbors from various ecosystems worldwide. For all genes coverage estimates and rarefaction analysis of unique sequences, however, showed that saturation was not reached, indicating that a large proportion of diversity remains unknown. Consequently, more large-scale analyses, including deep sequencing, is required to assess the diversity of these systems.

Acknowledgements

This work was funded by the Fund for Scientific Research – Flanders (project G.0146.12). Additional support was obtained from the Belgian Science Policy Office (project CCAMBIO).

The authors declare that they have no conflict of interest.

Fig. 1 Location of the Princess Elisabeth Station (base of flag post) in the Sør Rondane Mountains, East-Antarctica and, from left to right, photographs of the terrestrial samples used in this study (KP2, KP15, KP43 and KP53)

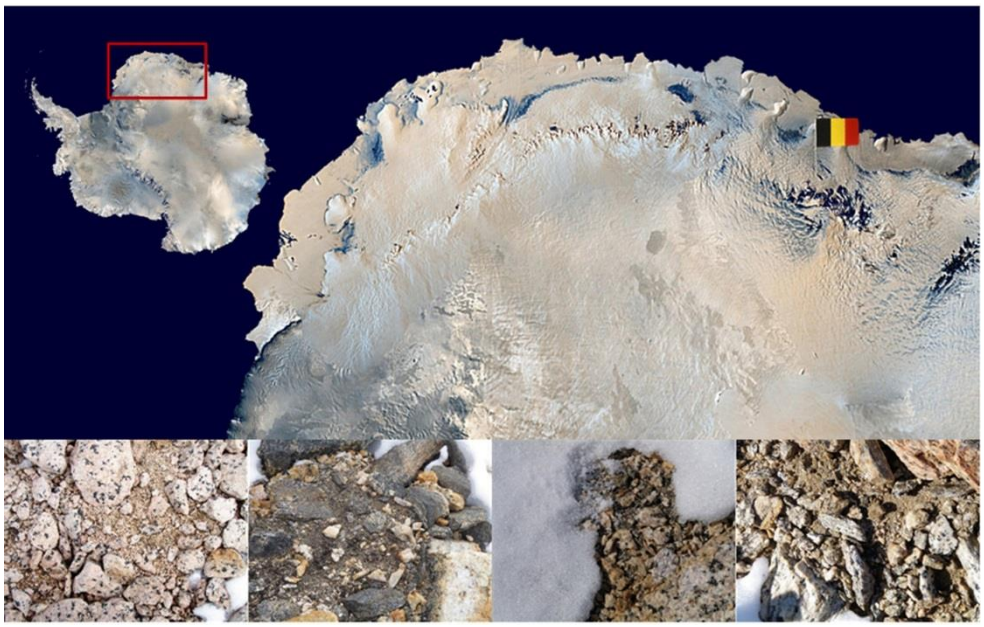


Fig. 2 Rarefaction curves based on OTUs grouping clones that have 95% and 100% similarity for nucleotide sequences, and 100% for protein sequences. Analysis was performed using the Vegan package in R

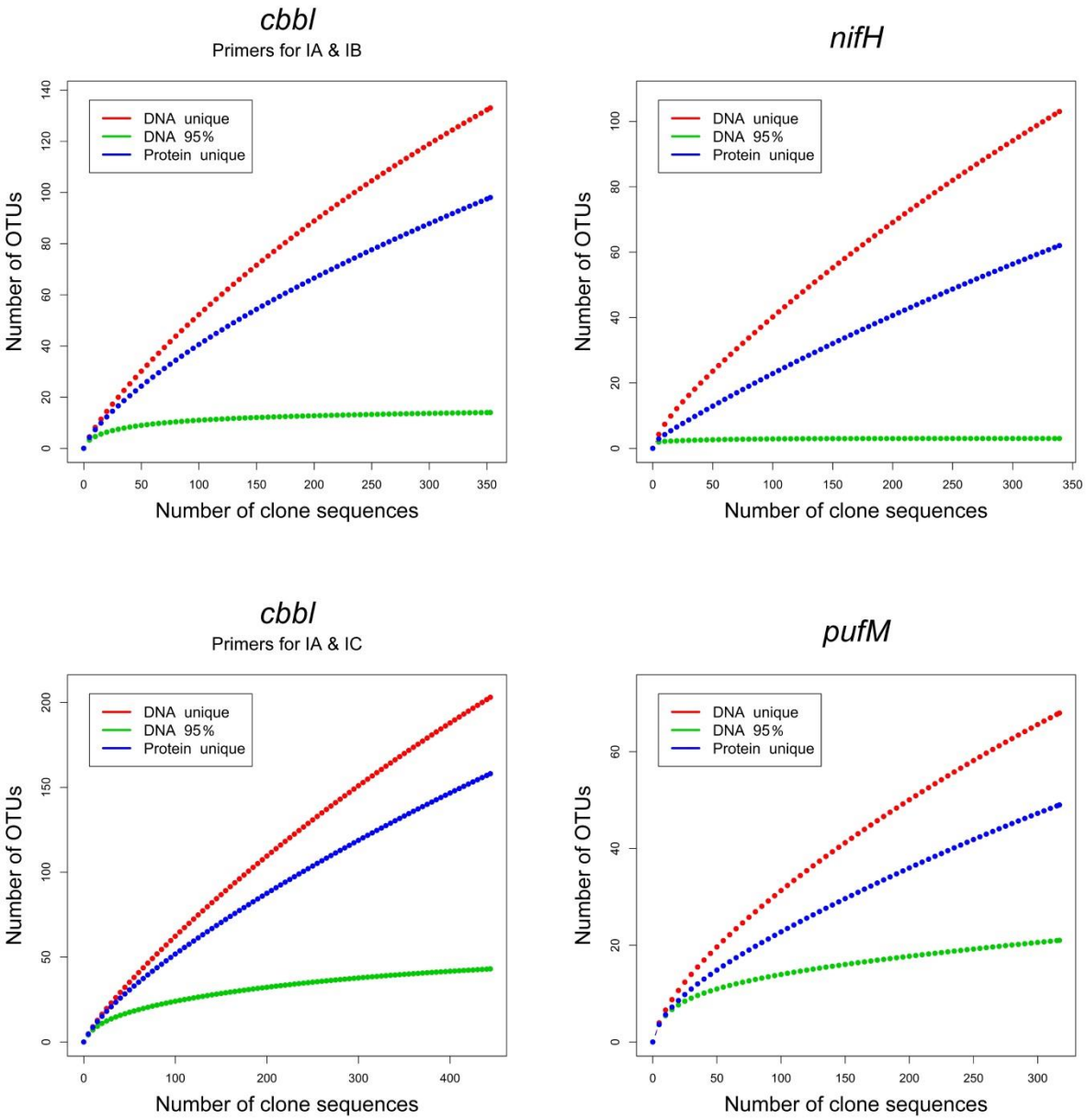


Fig. 3 Unweighted Venn diagram of gene protein ORUs/ONUs/OPUs, calculated using the VennDiagram package (http://cran.r-project.org/web/packages/VennDiagram/index.html) in R. Colored areas show the number of ORUs/ONUs/OPUs present in the terrestrial samples or shares between multiple samples

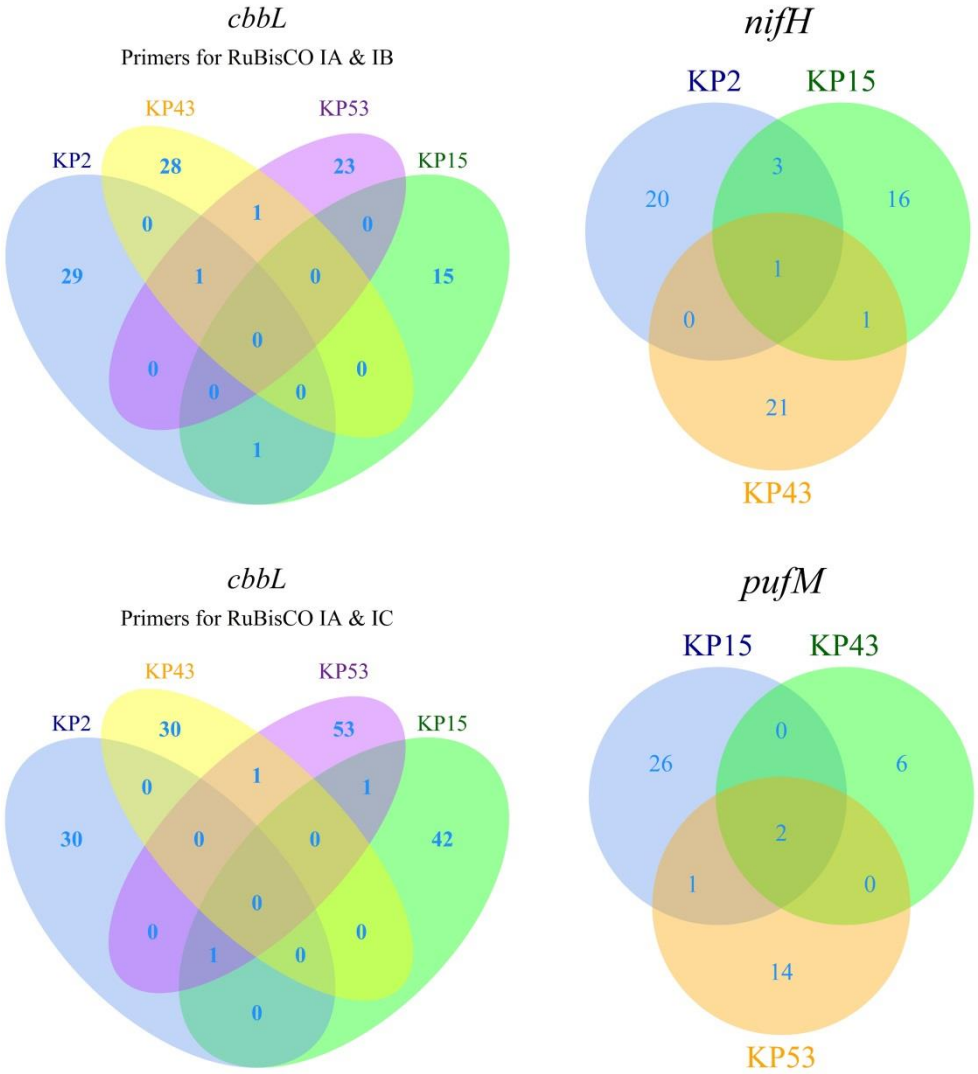
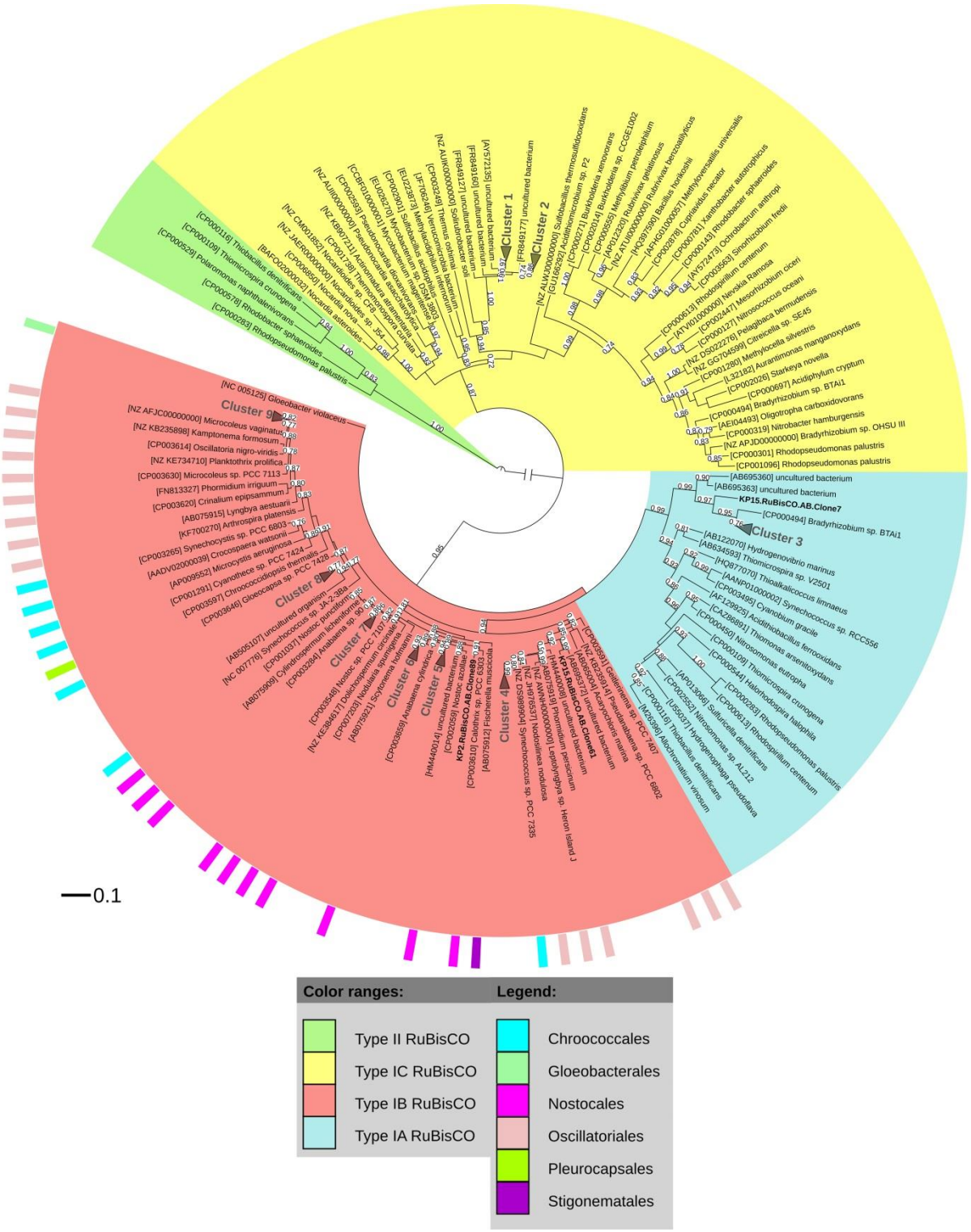


Fig. 4 Maximum likelihood phylogenetic tree (1.000 bootstrap replicates) of *cbbL* IA and IB amino acid sequences (267 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ORUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values < 70% are not shown. *cbbM* sequences of *Rhodopseudomonas palustris* BisB5, *Rhodobacter sphaeroides* ATCC 17029, *Polaromonas naphthalenivorans* CJ2, *Thiomicrospira crunogena* XCL-2 and *Thiobacillus denitrificans* ATCC 25259 (accession numbers CP000283, CP000578, CP000529, CP000109 and CP000116) were used as an outgroup. Different Cyanobacterial orders are labelled with a colored bar. Visual clades are displayed as simplified gray triangles


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
KP15.RuBisCO.AB.Clone7
KP2.RuBisCO.AB.Clone89
KP1.RuBisCO.AB.Clone61
0.1
Color ranges:
Type II RuBisCO
Type IC RuBisCO
Type IB RuBisCO
Type IA RuBisCO
Legend:
Chroococcales
Gloeobacterales
Nostocales
Oscillatoriales
Pleurocapsales
Stigonematales

484

485

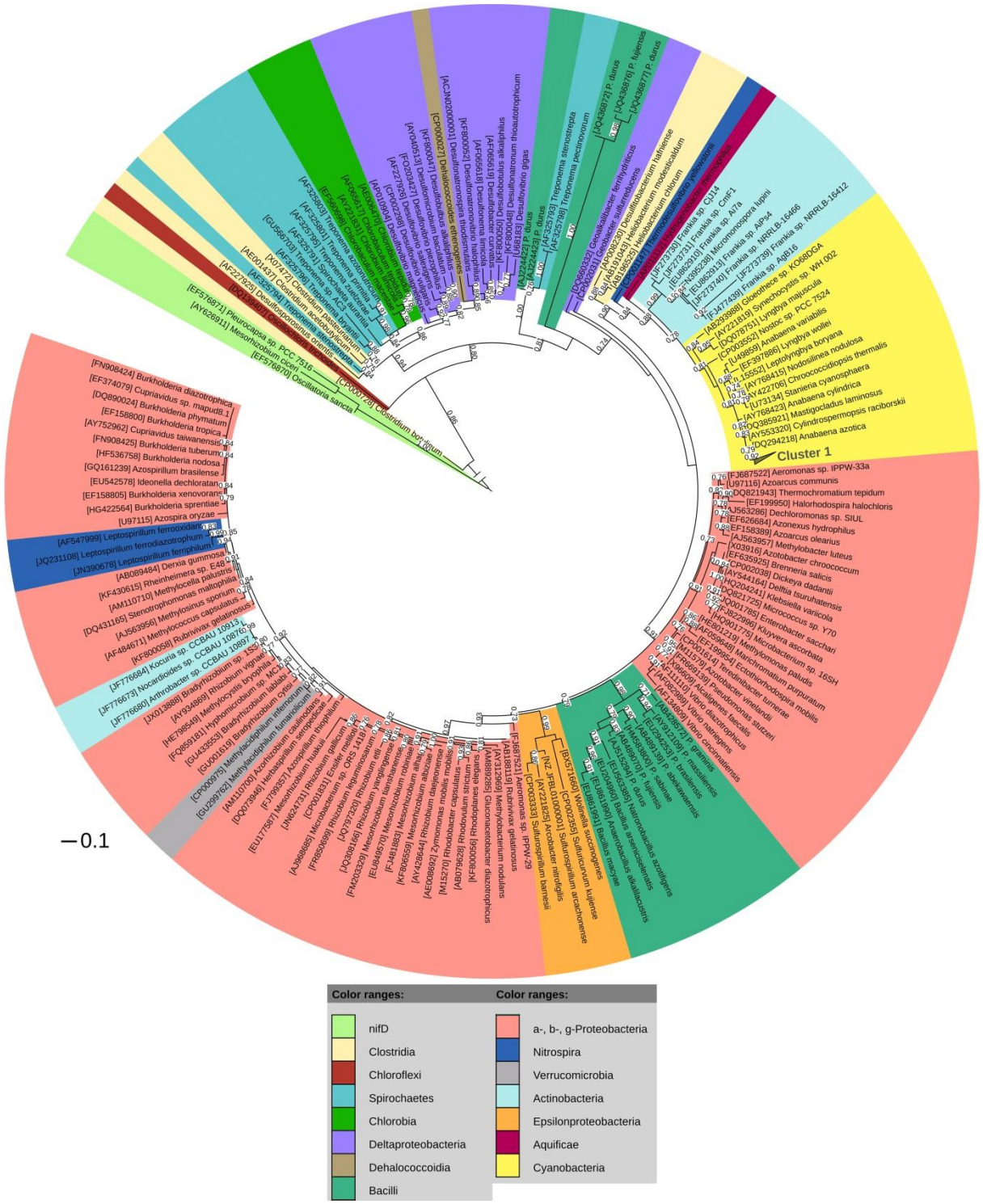
Fig. 5 Maximum likelihood phylogenetic tree (1.000 bootstrap replicates) of *cbbL* IA and IC amino acid sequences (154 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ORUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values < 70% are not shown. *cbbM* sequences of *Rhodopseudomonas palustris* BisB5, *Rhodobacter sphaeroides* ATCC 17029, *Polaromonas naphthalenivorans* CJ2 and *Thiomicrospira crunogena* XCL-2 (accession numbers CP000283, CP000578, CP000529 and CP000109) were used as an outgroup. Visual clades are displayed as gray triangles. Total branch lengths to the closest and the farthest leaf are used as sides of the triangle


0.1
Color ranges:
Type II RuBisCO
Type IB RuBisCO
Type IA RuBisCO
Type IC RuBisCO

495

496

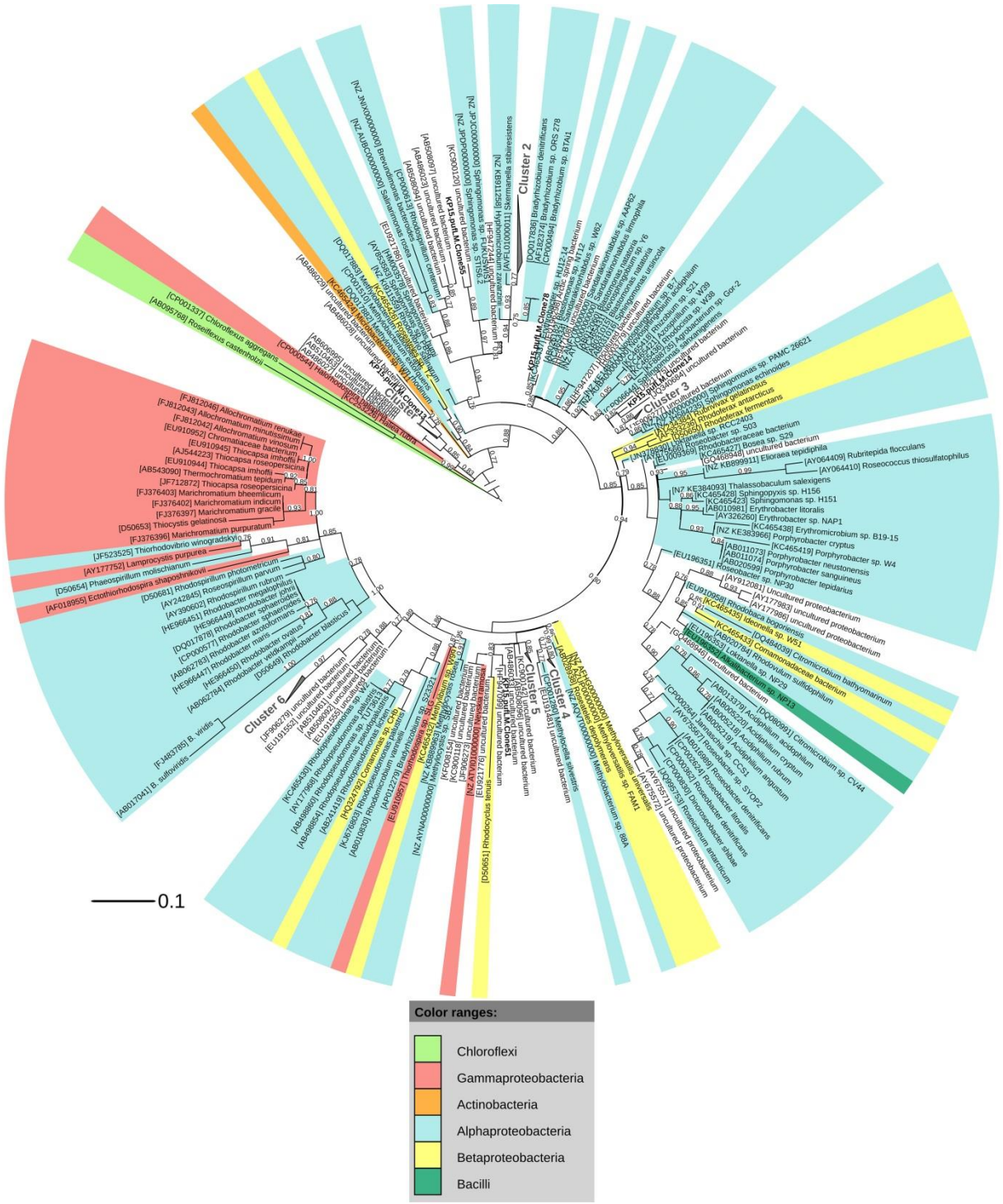
Fig. 6 Maximum likelihood phylogenetic tree (1.000 bootstrap replicates) of *nifH* amino acid sequences (106 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ONUs are labelled in bold. Accession numbers are listed (between brackets) next to each sequence name. Bootstrap values < 70% are not shown. *nifD* sequences of *Pleurocapsa* sp. PCC 7516 (EF576871), *Mesorhizobium ciceri* (AY626911) and *Oscillatoria sancta* PCC 7515 (EF576870) were used as an outgroup. Visual clades are displayed as gray triangles. Total branch lengths to the closest and the farthest leaf are used as sides of the triangle


Cluster 1
0.1
Color ranges:
nifD
Clostridia
Chloroflexi
Spirochaetes
Chlorobia
Deltaproteobacteria
Dehalococcoidia
Bacilli
Color ranges:
a-, b-, g-Proteobacteria
Nitrospira
Verrucomicrobia
Actinobacteria
Epsilonproteobacteria
Aquificae
Cyanobacteria

505

506

Fig. 7 Maximum likelihood phylogenetic tree (1.000 bootstrap replicates) of *pufM* amino acid sequences (58 AA). Scale bar indicates 10 substitutions per 100 amino acid positions. Clones are named by terrestrial sample_gene name_index number. Separate OPUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values < 70% are not shown. *Chloroflexus aggregans* and *Roseiflexus castenholzii* (accession numbers CP001337 and AB095768) were used as an outgroup. Visual clades are displayed as gray triangles. Total branch lengths to the closest and the farthest leaf are used as sides of the triangle


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
0.1
Color ranges:
Chloroflexi
Gammaproteobacteria
Actinobacteria
Alphaproteobacteria
Betaproteobacteria
Bacilli

514

Table 1 Parameters associated with analyzed samples

Sample	Sample Coordinates	Altitude (m)	Description of sample area	Conductivity (µS/cm)	pH	Water Content	TOC	Relative abundance Cyanobacteria[a]
KP2	71°57'28.6" S, 23°19'45.8" E	1320	Small gravel particles in between the rocks, moraine slope west	19	6.54	6.28%	0.08%	19.62%
KP15	71°56'45.8" S, 23°20'43.6" E	1366	Brown soil under lichen	33	5.57	3.38%	0.33%	51.85%
KP43	71°56’47.3” S, 23°20’44.6” E	1362	Brown soil with dark green fragments	520	6.22	0.91%	2.57%	0.04%
KP53	71°56’45.3” S, 23°20’42.4” E	1362	Grey soil on second part of the ridge	312	6.34	0.23%	0.21%	0.12%

[a] Data based on Illumina MiSeq 2x 300 bp paired-end sequencing of partial 16S rRNA genes (unpublished data)

Table 2 PCR Primers and conditions used for screening different genes

Gene	Target	Primer	Sequence 5'-3'	Final Concentration	Region	Amplicon size	Program
cbbL	RuBisCO IA & IB	RubIgF[a]	GAY TTC ACC AAR GAY GAY GA	0.4 µM	571-1382[i]	± 800 bp	95°C (3min); 3x 95°C (1min), 49°C (2min 15s), 72°C (2min 15s); 30x 95°C (35s), 49°C (1min 15s), 72°C (1min 15s); 72°C (7min)
		RubIgR[a]	TCR AAC TTG ATY TCY TTC CA	0.4 µM			
cbbL	RuBisCO IA & IC	K2f[b]	ACC AYC AAG CCS AAG CTS GG	0.2 µM	496-990[i]	492-495 bp	95°C (3min); 35x 95°C (1min), 62°C (1min), 72°C (1min 30s); 72°C (10min) [61]
		V2r[b]	GCC TTC SAG CTT GCC SAC CRC	0.2 µM			
cbbM	RuBisCO II	cbbM343F[c]	GGY AAY AAC CAR GGY ATG GG	0.1 µM	343-1126[c]	700-800 bp	95°C (3min); 30x 95°C (1min), 50°C (2min), 72°C (3min); 72°C (7min) [63]
		cbbM1126R[c]	CGY ARB GCR TTC ATR CCR CC	0.1 µM			
pufLM	AAP	pufLF[d]	CTK TTC GAC TTC TGG GTS GG	0.2 µM	64-1612[j]	± 1500 bp	94°C (3min); 30x 94°C (1min), 60°C (1min), 72°C (2min); 72°C (10min) [93]
		pufMR[e]	CCA TSG TCC AGC GCC AGA A	0.2 µM			
PR	*Flavo-bacteria*	PR-Flavo-F[f]	GAY TAY GTW GSW TTY ACD TTY TTT GTR GG	0.2 µM	58-572[k]	460 bp	94°C (3min); 15x 94°C (30s), 60°C (45s) (-0.5°C/cycle), 72°C (30s) ; 20x 94°C (30s), 52°C (45s), 72°C (30s) ; 72°C (10min) [85]
		PR-Flavo-R[f]	GCC CAW CCH ACW ARW ACR AAC CAR CAT A	0.2 µM			
PR	*Flavo-bacterium*	PR-Flavo-2F[f]	GGC TAT GAT GGC HGC WK	0.2 µM	93-586[k]	460 bp	
		PR-Flavo-2R[f]	CWA DWG GRT ARA TNG CCC A	0.2 µM			
PR	Universal	PR-1aF[g]	GAT CGA GCG NTA YRT HGA RTG G	1.87 µM	340-665[l]	± 335 bp	94°C (2min), 30x 94°C (30s), 52°C (30s), 72°C (30s) ; 72°C (7min) [94]
		PR-1aR[g]	GAT CGA GCR TAD ATN GCC CAN CC	1.87 µM			
nifH	Universal	PolF[h]	TGC GAY CCS AAR GCB GAC TC	0.3 µM	115-476[h]	360 bp	95°C (5min); 30x 94°C (30s), 53°C (1min), 72°C (40s); 72°C (5min) [95]
		PolR[h]	ATS GCC ATC ATY TCR CCG GA	0.3 µM			

[a] data from [59]

[b] data from [62]

[c] data from [63]

[d] data from [96]

[e] data from [97]

[f] data from [85]

[g] data from [94]

[h] data from [65]

[i] Based on the *cbbL* IA sequence of *Bradyrhizobium* sp. ORS278 (CU234118).

[j] Based on the *pufLM* sequence of *Sphingomonas sanxanigenens* DSM 19645 (CP006644).

[k] Based on the PR sequence of *Dokdonia* sp. PRO 95 (FJ627053).

[l] Based on the PR sequence of *Vibrio campbellii* BAA-1116 (FJ985782).

534 **Table 3** Overview of clone library characteristics. Diversity indices were calculated on the basis of

535 derived unique protein sequences

			Clone Library				
			KP2	**KP15**	**KP43**	**KP53**	**All**
cbbL	No. of clones		115	39	123	76	353
(Primers for IA/IB)	No. of OTUs (100% DNA)		44	17	43	30	133
	No. of OTUs (95% DNA)		6	5	5	4	14
	No. of ORUs (100% AA)		31	16	30	25	98
	evenness (H/H_{max})		0.652	0.896	0.478	0.708	0.728
	Chao1		52.86	20.20	130.00	82.00	276.75
	Coverage %		58.6%	79.2%	23.1%	30.5%	35.4%
	Bray-Curtis	KP15	0.97				
		KP43	0.99	1.00			
		KP53	0.99	1.00	0.98		
cbbL	No. of clones		98	108	122	116	444
(Primers for IA/IC)	No. of OTUs (100% DNA)		41	55	39	70	203
	No. of OTUs (95% DNA)		14	14	4	16	43
	No. of ORUs (100% AA)		31	44	31	56	158
	evenness (H/H_{max})		0.813	0.811	0.587	0.842	0.805
	Chao1		41.50	102.13	77.00	206.50	388.22
	Coverage %		74.7%	43.1%	40.3%	27.1%	40.7%
	Bray-Curtis	KP15	0.96				
		KP43	1.00	1.00			
		KP53	0.99	0.96	0.87		
nifH	No. of clones		115	119	105	/	339
	No. of OTUs (100% DNA)		33	31	41	/	103
	No. of OTUs (95% DNA)		2	3	2	/	3
	No. of ONUs (100% AA)		24	21	23	/	62
	evenness (H/H_{max})		0.415	0.392	0.459	/	0.466
	Chao1		87.33	47.25	50.20	/	50.48
	Coverage %		27.5%	44.4%	45.8%	/	37.7%
	Bray-Curtis	KP15	0.23				
		KP43	0.93	0.93			
pufM	No. of clones		/	98	98	121	317
	No. of OTUs (100% DNA)		/	37	16	19	68
	No. of OTUs (95% DNA)		/	16	4	6	21
	No. of OPUs (100% AA)		/	29	8	17	49
	evenness (H/H_{max})		/	0.814	0.496	0.525	0.612
	Chao1		/	46.14	8.75	56.00	98.60
	Coverage %		/	62.8%	91.4%	30.4%	49.7%
	Bray-Curtis	KP43		0.92			
		KP53		0.91	0.57		

Table 4 Composition of visual clusters and singletons and the origin of their closest neighbor, per clone library. Colors indicate *cbbL* subtype (Blue = IA, Red = IB, Yellow = IC), ENV refers to environmental clones

Gene	Cluster	No. of unique protein sequences	No. of Sequences				Closest neighbor	
			KP2	KP15	KP43	KP53	Accession no.	Origin
cbbL	1	38			108	43	AY572135	ENV; Germany; Agricultural soil
Primers for IA/IB	2	7				29	FR849177	ENV; China; Arid soil
	3	2		2			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*
	KP15.RuBisCO.AB.Clone7	1		2			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*
	4	4	9				NZ_DS989904	*Synechococcus* sp. PCC 7335; Mexico; Snail shell
	5	3			9		CP003659	*Anabaena cylindrica* PCC 7122; United Kingdom; Water
	6	11	16		5	4	CP003659	*Anabaena cylindrica* PCC 7122; United Kingdom; Water
	7	2	2	2	1		CP003284	*Anabaena* sp. 90; Finland; Lake water
	8	17	83				AB505107	ENV; Japan; Iron-rich snow
	9	11		32			AB695365[a]	ENV; Antarctica; Aquatic moss pillar
	KP15.RuBisCO.AB.Clone61	1		1			AB695372	ENV; Antarctica; Aquatic moss pillar
	KP2.RuBisCO.AB.Clone89	1	5				CP003610	*Calothrix* sp. PCC 6303; USA; Freshwater
cbbL	1	2	3				JQ964878	ENV; China; Agricultural soil
Primers for IA/IC	2	32			55	46	AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit
	3	17		45			KF523929	ENV; China; Soil
	4	7		15		2	NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil
	5	3	9				NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil
	6	5		15			JQ964837	ENV; China; Agricultural soil
	7	2			3		JAOB01000069	*Mycobacterium xenopi* 4042; Human lung
	8	4				6	JQ836503	ENV; China; Soil
	9	14	21			34	JQ836503	ENV; China; Soil
	10	2				2	NZ_JAER00000000	*Nocardioides* sp. J54; Chile; Atacama desert
	11	5		9		1	FO082843	*Nocardia cyriacigeorgica* GUH-2; Human
	12	2		3			NZ_BAFO00000000	*Nocardia asteroides* NBRC 15531
	13	11	36				GU166292	*Acidithiomicrobium* sp. P2; Pyrite enrichment culture
	14	2	3				HQ174601	ENV; China; Cropland soil
	15	6				12	HQ388736	ENV; China; Paddy soil
	16	17	4		63		CAUA01000016	*Nitrosospira* sp. APG3; USA; Freshwater lake sediment
	KP15.RuBisCO.AC.Clone47	1		1			JQ836457	ENV; China. Soil
	KP53.RuBisCO.AC.Clone141	1				1	JQ836514	ENV; China. Soil
	KP2.RuBisCO.AC.Clone100	1	2				NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil
	KP53.RuBisCO.AC.Clone4	1				3	AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit
	KP15.RuBisCO.AC.Clone109	1		2			AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit
	KP2.RuBisCO.AC.Clone102	1	1				JQ836516	ENV; China; Soil
	KP43.RuBisCO.AC.Clone116	1			1		JQ836516	ENV; China; Soil
	KP2.RuBisCO.AC.Clone84	1	1				NZ_CM001852	*Nocardioides* sp. CF8; USA; Soil
	KP53.RuBisCO.AC.Clone67	1				1	JQ964872	ENV; China; Soil
	KP15.RuBisCO.AC.Clone5	1		1			DQ149782	ENV; USA; Pine forest and agricultural soils

	KP2.RuBisCO.AC.Clone103	1	4				JQ964901	ENV; China; Soil
	KP15.RuBisCO.AC.Clone8	1		2			NZ_ATUI00000000	*Rubrivivax benzoatilyticus* JA2; India; Paddy soil
	KP15.RuBisCO.AC.Clone148	1		2			KJ720517	ENV; China; Soil
	KP15.RuBisCO.AC.Clone104	1		5		3	HQ174649	ENV; China; Cropland soil
	KP15.RuBisCO.AC.Clone152	1		1			HQ174649	ENV; China; Cropland soil
	KP15.RuBisCO.AC.Clone52	1		1			HQ174649	ENV; China; Cropland soil
	KP53.RuBisCO.AC.Clone92	1				1	KJ720371	ENV; China; Soil
	KP53.RuBisCO.AC.Clone47	1				1	KJ720371	ENV; China; Soil
	KP53.RuBisCO.AC.Clone25	1				1	AY422900	ENV; Hawaii; 300-yr old forest soil volcanic deposit
	KP2.RuBisCO.AC.Clone77	1	1				HQ997091	ENV; China; Cropland soil
	KP2.RuBisCO.AC.Clone117	1	1				KJ720363	ENV; China; Soil
	KP2.RuBisCO.AC.Clone1	1	2				KJ720363	ENV; China; Soil
	KP53.RuBisCO.AC.Clone83	1				1	HQ997138	ENV; China; Cropland soil
	KP15.RuBisCO.AC.Clone103	1	7	4		1	KJ720441	ENV; China; Soil
	KP15.RuBisCO.AC.Clone18	1		1			JQ836500	ENV; China; Soil
	KP2.RuBisCO.AB.Clone110	1	3				DQ149787	ENV; USA; Pine forest and agricultural soils
	KP15.RuBisCO.AC.Clone113	1		1			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*
nifh	1[b]	60	110	118	105	/		*Nostocaceae* Cyanobacteria
pufM	1	10	/	5	68	22	AB486028	ENV; China; Paddy soil
	2	13	/	24		24	AVFL01000011	*Skermanella stibiiresistens* SB22; China; Iron mine soil
	3	4	/	24			JF906277	ENV; Germany; Freshwater lake
	4	2	/	17			CP001280	*Methylocella silvestris* BL2; Germany; Acidic forest cambisol
	5	4	/	14			KC900142	ENV; Antarctica; Great Wall Cove
	6	11	/	8	30	75	JF906279	ENV; Germany; Freshwater lake
	KP15.pufLM.Clone55	1	/	2			KC900120	ENV; Antarctica; Great Wall Cove
	KP15.pufLM.Clone78	1	/	1			KC465437	*Erythrobacter* sp. HU12-14; Mongolia; Lake Hulunhu
	KP15.pufLM.Clone51	1	/	1			HF947099	ENV; China; Lake Taihu
	KP15.pufLM.Clone14	1	/	1			JQ340684	ENV; Pacific Ocean
	KP15.pufLM.Clone13	1	/	1			AB486028	ENV; China; Paddy soil

[a] The environmental sequence with accession number AB695365 is enclosed in the cluster.

[b] The following reference sequences are enclosed in the cluster: L23514 (*Nostoc commune*), CP001037 (*Nostoc punctiforme* PCC 73102), KC243672 (*Nostoc* sp. LEGE 06106), U04054 (*Nostoc muscorum*), CP007203 (*Nodularia spumigena* CCY9414), CP003642 (*Cylindrospermum stagnale* PCC 7417), CP003943 (*Calothrix* sp. PCC 7507), and environmental sequences DQ995890, EU915059, AY819602, HM140755, DQ140640, GU196863, KC667460 and KC140445.

Supplementary Material

Analysis of *cbbL*, *nifH* and *pufLM* in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of auto- and phototrophic bacteria

Guillaume Tahon, Bjorn Tytgat, Pieter Stragier, Anne Willems*

***Correspondence:** Anne Willems, Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, Ghent, Belgium, Anne.Willems@UGent.be

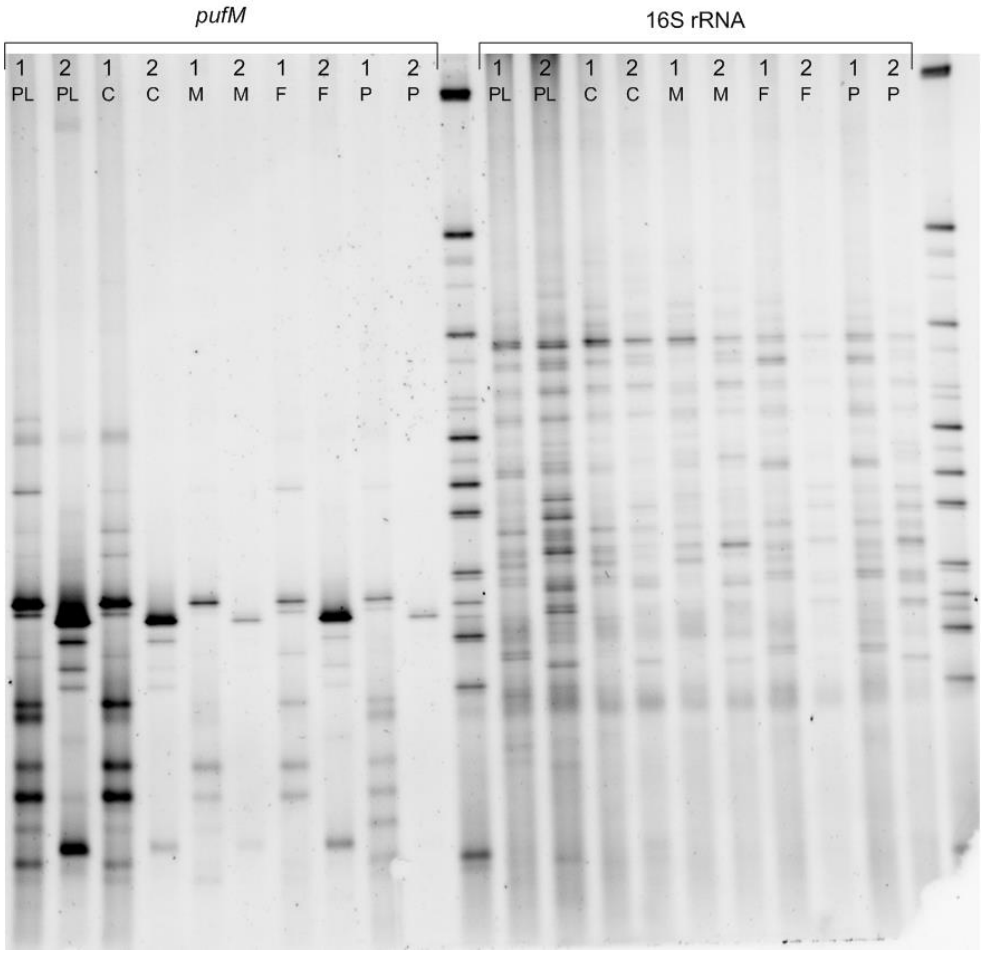


Fig. S1 Comparison of genetic fingerprints (DGGE) of two terrestrial samples collected near the Princess Elisabeth Station, by use of five different DNA extraction methods. Partial 16S (V1-V3 region) was amplified using primers V3-357F_GC and V3-518R (Temmerman et al., 2003). Partial *pufM* gene fragments were amplified using primers pufM557F (Achenbach et al., 2001) and

GC_WAWR (Yutin et al., 2008). Gene fragments retrieved were visualized using DGGE. Lanes designated as 1 and 2 correspond with the two terrestrial samples tested. Different DNA extraction methods used are labeled as PL (PowerLyzer® PowerSoil® DNA Isolation Kit), C (CTAB), M (Muyzer), F (FastDNA™ SPIN Kit for Soil) and P (Pitcher). In total, seven different DNA extraction methods were tested: CTAB, Pitcher, Muyzer, PowerLyzer® PowerSoil® DNA Isolation Kit, FastDNA™ SPIN Kit for Soil, PowerSoil® DNA Isolation Kit and UltraClean Soil DNA Isolation Kit. For all methods, protocols were tested with and without modifications (e.g. the influence of bead bating or an extra lysis step), to optimize DNA yield and diversity. Partial *pufM* and 16S rRNA were amplified as described above and visualized using DGGE. The resulting best five DNA extraction methods were compared in the figure

Supplementary Material

Analysis of *cbbL, nifH* and *pufLM* in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of auto- and phototrophic bacteria

Guillaume Tahon, Bjorn Tytgat, Pieter Stragier, Anne Willems*

***Correspondence:** Anne Willems, Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, Ghent, Belgium, Anne.Willems@UGent.be

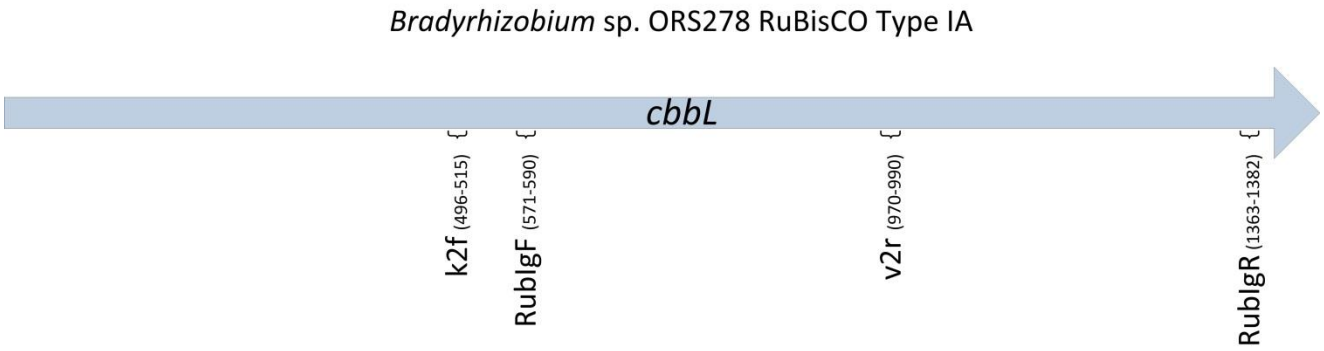


Fig. S2 Schematic illustration of the *cbbL* gene of *Bradyrhizobium* sp. ORS278 (accession number CU234118). Curly brackets indicate the positions of forward and reverse primers used in this study (Table 2), to target the different subtypes of RuBisCO type I. Positions of the primers are assigned according to the DNA sequence of the *cbbL* gene of *Bradyrhizobium* sp. ORS278 and given next to the primer name, between brackets

584 **References**

585 1. Wynn-Williams DD (1990) Microbial Colonization Processes In Antarctic Fellfield Soils : An
586 Experimental Overview (Eleventh Symposium on Polar Biology). Proceedings of the NIPR
587 Symposium on Polar Biology 3: 164-178.

588 2. Convey P, Gibson JAE, Hillenbrand C-D, Hodgson DA, Pugh PJA, Smellie JL, Stevens MI (2008)
589 Antarctic terrestrial life – challenging the history of the frozen continent? Biological Reviews
590 83: 103-117. doi: 10.1111/j.1469-185X.2008.00034.x

591 3. Matsuoka N, Thomachot CE, Oguchi CT, Hatta T, Abe M, Matsuzaki H (2006) Quaternary
592 bedrock erosion and landscape evolution in the Sor Rondane Mountains, East Antarctica:
593 Reevaluating rates and processes. Geomorphology 81: 408-420. doi:
594 10.1016/j.geomorph.2006.05.005

595 4. Convey P, Stevens MI (2007) Antarctic Biodiversity. Science 317: 1877-1878. doi:
596 10.1126/science.1147261

597 5. Greene DM, Holtom A (1971) Studies in Colobanthus quitensis (Kunth) Bartl. & Deschampsia
598 antarctica Desv. III. Distribution, habitats and perfomance in the Antarctic botanical zone.
599 British Antarctic Survey Bulletin 26: 1-29.

600 6. Michaud AB, Šabacká M, Priscu JC (2012) Cyanobacterial diversity across landscape units in a
601 polar desert: Taylor Valley, Antarctica. FEMS Microbiology Ecology: n/a-n/a. doi:
602 10.1111/j.1574-6941.2012.01297.x

603 7. Cowan DA, Sohm JA, Makhalanyane TP, Capone DG, Green TGA, Cary SC, Tuffin IM (2011)
604 Hypolithic communities: important nitrogen sources in Antarctic desert soils. Environmental
605 Microbiology Reports 3: 581-586. doi: 10.1111/j.1758-2229.2011.00266.x

606 8. Moodley K (2004) Microbial Diversity of Antarctic Dry Valley Mineral Soil. University of the
607 Western Cape

608 9. Yergeau E, Newsham KK, Pearce DA, Kowalchuk GA (2007) Patterns of bacterial diversity
609 across a range of Antarctic terrestrial habitats. Environmental Microbiology 9: 2670-2682. doi:
610 10.1111/j.1462-2920.2007.01379.x

611 10. Yergeau E, Bokhorst S, Huiskes AHL, Boschker HTS, Aerts R, Kowalchuk GA (2007) Size and
612 structure of bacterial, fungal and nematode communities along an Antarctic environmental
613 gradient. Fems Microbiology Ecology 59: 436-451. doi: 10.1111/j.1574-6941.2006.00200.x

614 11. Tabita FR (1999) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different
615 perspective. Photosynthesis Research 60: 1-28. doi: 10.1023/a:1006211417981

616 12. Kellermann C (2009) Autotrophy in Groundwater Ecosystems. Dissertation, LMU Munich,

617 13. Tourova TP, Spiridonova EM (2009) Phylogeny and Evolution of the Ribulose 1,5-
618 Bisphosphate Carboxylase/Oxygenase Genes in Prokaryotes. Molecular Biology 43: 713-728.
619 doi: 10.1134/s0026893309050033

620 14. Badger MR, Bek EJ (2008) Multiple Rubisco forms in proteobacteria: their functional
621 significance in relation to CO2 acquisition by the CBB cycle. Journal of Experimental Botany
622 59: 1525-1541. doi: 10.1093/jxb/erm297

623 15. Khadem AF, Pol A, Wieczorek A, Mohammadi SS, Francoijs K-J, Stunnenberg HG, Jetten MSM,
624 Op den Camp HJM (2011) Autotrophic Methanotrophy in Verrucomicrobia:
625 Methylacidiphilum fumariolicumSolV Uses the Calvin-Benson-Bassham Cycle for Carbon
626 Dioxide Fixation. Journal of Bacteriology 193: 4438-4446. doi: 10.1128/jb.00407-11

16. Caldwell PE, MacLean MR, Norris PR (2007) Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in Sulfobacillus species. Microbiology 153: 2231-2240. doi: 10.1099/mic.0.2007/006262-0

17. Grostern A, Alvarez-Cohen L (2013) RubisCO-based CO2 fixation and C1 metabolism in the actinobacterium Pseudonocardia dioxanivorans CB1190. Environmental Microbiology 15: 3040-3053. doi: 10.1111/1462-2920.12144

18. Park SW, Hwang EH, Jang HS, Lee JH, Kang BS, Oh JI, Kim YM (2009) Presence of duplicate genes encoding a phylogenetically new subgroup of form I ribulose 1,5-bisphosphate carboxylase/oxygenase in Mycobacterium sp. strain JC1 DSM 3803. Research in Microbiology 160: 159-165. doi: 10.1016/j.resmic.2008.12.002

19. Van Der Wielen PWJJ (2006) Diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes in the MgCl2-dominated deep hypersaline anoxic basin discovery. FEMS Microbiology Letters 259: 326-331. doi: 10.1111/j.1574-6968.2006.00284.x

20. Selesi D, Schmid M, Hartmann A (2005) Diversity of Green-Like and Red-Like Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes (cbbL) in Differently Managed Agricultural Soils. Applied and Environmental Microbiology 71: 175-184. doi: 10.1128/aem.71.1.175-184.2005

21. Nakai R, Abe T, Baba T, Imura S, Kagoshima H, Kanda H, Kohara Y, Koi A, Niki H, Yanagihara K, Naganuma T (2012) Diversity of RuBisCO gene responsible for CO₂ fixation in an Antarctic moss pillar. Polar Biology 35: 1641-1650. doi: 10.1007/s00300-012-1204-5

22. Hanson TE, Tabita FR (2001) A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. Proceedings of the National Academy of Sciences of the United States of America 98: 4397-4402. doi: 10.1073/pnas.081610398

23. Howard KS, McLean PA, Hansen FB, Lemley PV, Koblan KS, Ormejohnson WH (1986) *Klebsiella pneumoniae nifM* Gene Product Is Required For Stabilization And Activation Of Nitrogenase Iron Protein In *Escherichia coli*. Journal of Biological Chemistry 261: 772-778.

24. Fogg GE, Stewart WDP (1968) In Situ Determination of Biological Nitrogen Fixation in Antarctica. British Antarctic Survey Bulletin 15: 36-46.

25. Fritsen CH, Grue AM, Priscu JC (2000) Distribution of organic carbon and nitrogen in surface soils in the McMurdo Dry Valleys, Antarctica. Polar Biology 23: 121-128. doi: 10.1007/s003000050017

26. Zehr JP, Jenkins BD, Short SM, Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental Microbiology 5: 539-554. doi: 10.1046/j.1462-2920.2003.00451.x

27. Yeager CM, Kornosky JL, Housman DC, Grote EE, Belnap J, Kuske CR (2004) Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado plateau and Chihuahuan desert. Applied and Environmental Microbiology 70: 973-983. doi: 10.1128/aem.70.2.973-983.2004

28. Palovaara J, Akram N, Baltar F, Bunse C, Forsberg J, Pedros-Alio C, Gonzalez JM, Pinhassi J (2014) Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. Proceedings of the National Academy of Sciences of the United States of America 111: E3650-E3658. doi: 10.1073/pnas.1402617111

29. Beja O, Lanyi JK (2014) Nature's toolkit for microbial rhodopsin ion pumps. Proceedings of the National Academy of Sciences of the United States of America 111: 6538-6539. doi: 10.1073/pnas.1405093111

30. Bamann C, Bamberg E, Wachtveitl J, Glaubitz C (2014) Proteorhodopsin. Biochimica Et Biophysica Acta-Bioenergetics 1837: 614-625. doi: 10.1016/j.bbabio.2013.09.010

31. Hauruseu D, Koblížek M (2012) Influence of Light on Carbon Utilization in Aerobic Anoxygenic Phototrophs. Applied and Environmental Microbiology 78: 7414-7419. doi: 10.1128/aem.01747-12

32. Koblizek M, Beja O, Bidigare RR, Christensen S, Benitez-Nelson B, Vetriani C, Kolber MK, Falkowski PG, Kolber ZS (2003) Isolation and characterization of Erythrobacter sp strains from the upper ocean. Archives of Microbiology 180: 327-338. doi: 10.1007/s00203-003-0596-6

33. Alberti M, Burke D, Hearst J (1995) Structure and Sequence of the Photosynthesis Gene Cluster. In: Blankenship, R, Madigan, M, Bauer, C (eds.) Anoxygenic Photosynthetic Bacteria. Springer Netherlands, pp. 1083-1106

34. Jiang H, Deng S, Huang Q, Dong H, Yu B (2010) Response of Aerobic Anoxygenic Phototrophic Bacterial Diversity to Environment Conditions in Saline Lakes and Daotang River on the Tibetan Plateau, NW China. Geomicrobiology Journal 27: 400-408. doi: 10.1080/01490450903480269

35. Salka I, Cuperova Z, Masin M, Koblizek M, Grossart H-P (2011) Rhodoferax-related pufM gene cluster dominates the aerobic anoxygenic phototrophic communities in German freshwater lakes. Environmental Microbiology 13: 2865-2875. doi: 10.1111/j.1462-2920.2011.02562.x

36. Perreault NN, Greer CW, Andersen DT, Tille S, Lacrampe-Couloume G, Lollar BS, Whyte LG (2008) Heterotrophic and Autotrophic Microbial Populations in Cold Perennial Springs of the High Arctic. Applied and Environmental Microbiology 74: 6898-6907. doi: 10.1128/aem.00359-08

37. Achenbach LA, Carey J, Madigan MT (2001) Photosynthetic and Phylogenetic Primers for Detection of Anoxygenic Phototrophs in Natural Environments. Applied and Environmental Microbiology 67: 2922-2926. doi: 10.1128/aem.67.7.2922-2926.2001

38. Cottrell MT, Kirchman DL (2009) Photoheterotrophic Microbes in the Arctic Ocean in Summer and Winter. Applied and Environmental Microbiology 75: 4958-4966. doi: 10.1128/aem.00117-09

39. Jeanthon C, Boeuf D, Dahan O, Le Gall F, Garczarek L, Bendif EM, Lehours A-C (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. Biogeosciences 8: 1955-1970. doi: 10.5194/bg-8-1955-2011

40. Lehours A-C, Cottrell MT, Dahan O, Kirchman DL, Jeanthon C (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. FEMS Microbiology Ecology 74: 397-409. doi: 10.1111/j.1574-6941.2010.00954.x

41. Ritchie AE, Johnson ZI (2012) Abundance and Genetic Diversity of Aerobic Anoxygenic Phototrophic Bacteria of Coastal Regions of the Pacific Ocean. Applied and Environmental Microbiology 78: 2858-2866. doi: 10.1128/aem.06268-11

42. Yutin N, Suzuki MT, Béjà O (2005) Novel Primers Reveal Wider Diversity among Marine Aerobic Anoxygenic Phototrophs. Applied and Environmental Microbiology 71: 8958-8962. doi: 10.1128/aem.71.12.8958-8962.2005

43. Karr EA, Sattley WM, Jung DO, Madigan MT, Achenbach LA (2003) Remarkable Diversity of Phototrophic Purple Bacteria in a Permanently Frozen Antarctic Lake. Applied and Environmental Microbiology 69: 4910-4914. doi: 10.1128/aem.69.8.4910-4914.2003

44. Jiang H, Dong H, Yu B, Lv G, Deng S, Wu Y, Dai M, Jiao N (2009) Abundance and diversity of aerobic anoxygenic phototrophic bacteria in saline lakes on the Tibetan plateau. FEMS Microbiology Ecology 67: 268-278. doi: 10.1111/j.1574-6941.2008.00616.x

45. Koh EY, Phua W, Ryan KG (2011) Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater. Environmental Microbiology Reports 3: 710-716. doi: 10.1111/j.1758-2229.2011.00286.x

46. Tytgat B, Verleyen E, Obbels D, Peeters K, De Wever A, D'Hondt S, De Meyer T, Van Criekinge W, Vyverman W, Willems A (2014) Bacterial Diversity Assessment in Antarctic Terrestrial and Aquatic Microbial Mats: A Comparison between Bidirectional Pyrosequencing and Cultivation. Plos One 9. doi: 10.1371/journal.pone.0097564

47. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Research 40: D115-D122. doi: 10.1093/nar/gkr1044

48. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729. doi: 10.1093/molbev/mst197

49. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75: 7537-7541. doi: 10.1128/aem.01541-09

50. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194-2200. doi: 10.1093/bioinformatics/btr381

51. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soeding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7. doi: 10.1038/msb.2011.75

52. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Research 38: W695-W699. doi: 10.1093/nar/gkq313

53. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490. doi: 10.1371/journal.pone.0009490

54. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127-128. doi: 10.1093/bioinformatics/btl529

55. Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Research 39: W475-W478. doi: 10.1093/nar/gkr201

56. Dixon P (2003) VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14: 927-930. doi: 10.1111/j.1654-1103.2003.tb02228.x

760 57. Elsaied H, Naganuma T (2001) Phylogenetic Diversity of Ribulose-1,5-Bisphosphate
761 Carboxylase/Oxygenase Large-Subunit Genes from Deep-Sea Microorganisms. Applied and
762 Environmental Microbiology 67: 1751-1765. doi: 10.1128/aem.67.4.1751-1765.2001

763 58. Kovaleva OL, Tourova TP, Muyzer G, Kolganova TV, Sorokin DY (2011) Diversity of RuBisCO
764 and ATP citrate lyase genes in soda lake sediments. Fems Microbiology Ecology 75: 37-47. doi:
765 10.1111/j.1574-6941.2010.00996.x

766 59. Spiridonova EM, Berg IA, Kolganova TV, Ivanovsky RN, Kuznetsov BB, Tourova TP (2004) An
767 Oligonucleotide Primer System for Amplification of the Ribulose-1,5-Bisphosphate
768 Carboxylase/Oxygenase Genes of Bacteria of Various Taxonomic Groups. Microbiology 73:
769 316-325. doi: 10.1023/b:mici.0000032243.93917.30

770 60. Elsaied H, Kimura H, Naganuma T (2007) Composition of archaeal, bacterial, and eukaryal
771 RuBisCO genotypes in three Western Pacific arc hydrothermal vent systems. Extremophiles
772 11: 191-202. doi: 10.1007/s00792-006-0025-2

773 61. Yuan H, Ge T, Chen C, O'Donnell AG, Wu J (2012) Significant Role for Microbial Autotrophy in
774 the Sequestration of Soil Carbon. Applied and Environmental Microbiology 78: 2328-2336.
775 doi: 10.1128/aem.06881-11

776 62. Nanba K, King GM, Dunfield K (2004) Analysis of Facultative Lithotroph Distribution and
777 Diversity on Volcanic Deposits by Use of the Large Subunit of Ribulose 1,5-Bisphosphate
778 Carboxylase/Oxygenase. Applied and Environmental Microbiology 70: 2245-2253. doi:
779 10.1128/aem.70.4.2245-2253.2004

780 63. Kato S, Nakawake M, Ohkuma M, Yamagishi A (2012) Distribution and phylogenetic diversity
781 of cbbM genes encoding RubisCO form II in a deep-sea hydrothermal field revealed by newly
782 designed PCR primers. Extremophiles 16: 277-283. doi: 10.1007/s00792-011-0428-6

783 64. Kojima H, Fukuhara H, Fukui M (2009) Community structure of microorganisms associated
784 with reddish-brown iron-rich snow. Systematic and Applied Microbiology 32: 429-437. doi:
785 10.1016/j.syapm.2009.06.003

786 65. Poly F, Monrozier LJ, Bally R (2001) Improvement in the RFLP procedure for studying the
787 diversity of nifH genes in communities of nitrogen fixers in soil. Research in Microbiology 152:
788 95-103. doi: 10.1016/s0923-2508(00)01172-4

789 66. Gaby JC, Buckley DH (2012) A Comprehensive Evaluation of PCR Primers to Amplify the nifH
790 Gene of Nitrogenase. Plos One 7. doi: 10.1371/journal.pone.0042149

791 67. Feng Y, Lin X, Wang Y, Zhang J, Mao T, Yin R, Zhu J (2009) Free-air CO2 enrichment (FACE)
792 enhances the biodiversity of purple phototrophic bacteria in flooded paddy soil. Plant and
793 Soil 324: 317-328. doi: 10.1007/s11104-009-9959-3

794 68. Feng Y, Lin X, Yu Y, Zhu J (2011b) Elevated Ground-Level O-3 Changes the Diversity of
795 Anoxygenic Purple Phototrophic Bacteria in Paddy Field. Microbial Ecology 62: 789-799. doi:
796 10.1007/s00248-011-9895-7

797 69. Feng Y, Lin X, Mao T, Zhu J (2011a) Diversity of aerobic anoxygenic phototrophic bacteria in
798 paddy soil and their response to elevated atmospheric CO2. Microbial Biotechnology 4: 74-81.
799 doi: 10.1111/j.1751-7915.2010.00211.x

800 70. Kong W, Dolhi JM, Chiuchiolo A, Priscu J, Morgan-Kiss RM (2012) Evidence of form II RubisCO
801 (cbbM) in a perennially ice-covered Antarctic lake. Fems Microbiology Ecology 82: 491-500.
802 doi: 10.1111/j.1574-6941.2012.01431.x

71. Kong W, Ream DC, Priscu JC, Morgan-Kiss RM (2012) Diversity and expression of RubisCO genes in a perennially ice-covered Antarctic lake during the polar night transition. Applied and Environmental Microbiology. doi: 10.1128/aem.00029-12

72. Tourova TP, Kovaleva OL, Sorokin DY, Muyzer G (2010) Ribulose-1,5-bisphosphate carboxylase/oxygenase genes as a functional marker for chemolithoautotrophic halophilic sulfur-oxidizing bacteria in hypersaline habitats. Microbiology 156: 2016-2025. doi: 10.1099/mic.0.034603-0

73. Tourova TP, Kovaleva OL, Bumazhkin BK, Patutina EO, Kuznetsov BB, Bryantseva IA, Gorlenko VM, Sorokin DY (2011) Application of ribulose-1,5-bisphosphate carboxylase/oxygenase genes as molecular markers for assessment of the diversity of autotrophic microbial communities inhabiting the upper sediment horizons of the saline and soda lakes of the Kulunda Steppe. Microbiology 80: 812-825. doi: 10.1134/S0026261711060221

74. Yousuf B, Keshri J, Mishra A, Jha B (2012) Application of targeted metagenomics to explore abundance and diversity of CO2-fixing bacterial community using cbbL gene from the rhizosphere of Arachis hypogaea. Gene 506: 18-24. doi: 10.1016/j.gene.2012.06.083

75. Boyd ES, Hamilton TL, Havig JR, Skidmore ML, Shock EL (2014) Chemolithotrophic Primary Production in a Subglacial Ecosystem. Applied and Environmental Microbiology 80: 6146-6153. doi: 10.1128/aem.01956-14

76. Delwiche CF, Palmer JD (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Molecular Biology and Evolution 13: 873-882.

77. Díez B, Bergman B, Pedrós-Alió C, Antó M, Snoeijs P (2012) High cyanobacterial nifH gene diversity in Arctic seawater and sea ice brine. Environmental Microbiology Reports 4: 360-366. doi: 10.1111/j.1758-2229.2012.00343.x

78. Olson JB, Steppe TF, Litaker RW, Paerl HW (1998) N<inf>2</inf>-Fixing Microbial Consortia Associated with the Ice Cover of Lake Bonney, Antarctica. Microbial Ecology 36: 231-238. doi: 10.1007/s002489900110

79. Dodds WK, Gudder DA, Mollenhauer D (1995) The Ecology Of Nostoc. Journal of Phycology 31: 2-18. doi: 10.1111/j.0022-3646.1995.00002.x

80. Hawes I, Howardwilliams C, Vincent WF (1992) Desiccation And Recovery Of Antarctic Cyanobacterial Mats. Polar Biology 12: 587-594.

81. Teixeira LCRS, Peixoto RS, Cury JC, Sul WJ, Pellizari VH, Tiedje J, Rosado AS (2010) Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. ISME J 4: 989-1001. doi: 10.1038/ismej.2010.35

82. Jung JY, Choi AR, Lee YK, Lee HK, Jung K-H (2008) Spectroscopic and photochemical analysis of proteorhodopsin variants from the surface of the Arctic Ocean. FEBS Letters 582: 1679-1684. doi: 10.1016/j.febslet.2008.04.025

83. Atamna-Ismaeel N, Sabehi G, Sharon I, Witzel K-P, Labrenz M, Jurgens K, Barkay T, Stomp M, Huisman J, Beja O (2008) Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. ISME J 2: 656-662.

84. Koh EY, Atamna-Ismaeel N, Martin A, Cowie ROM, Beja O, Davy SK, Maas EW, Ryan KG (2010) Proteorhodopsin-Bearing Bacteria in Antarctic Sea Ice. Applied and Environmental Microbiology 76: 5918-5925. doi: 10.1128/aem.00562-10

85. Yoshizawa S, Kawanabe A, Ito H, Kandori H, Kogure K (2012) Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. Environmental Microbiology 14: 1240-1248. doi: 10.1111/j.1462-2920.2012.02702.x

86. Sabehi G, Loy A, Jung K-H, Partha R, Spudich JL, Isaacson T, Hirschberg J, Wagner M, Béjà O (2005) New Insights into Metabolic Properties of Marine Bacteria Encoding Proteorhodopsins. PLoS Biol 3: e273. doi: 10.1371/journal.pbio.0030273

87. Tank M, Thiel V, Imhoff JF (2009) Phylogenetic relationship of phototrophic purple sulfur bacteria according to pufL and pufM genes. International Microbiology 12: 175-185. doi: 10.2436/20.1501.01.96

88. Feng Y, Grogan P, Caporaso JG, Zhang H, Lin X, Knight R, Chu H (2014) pH is a good predictor of the distribution of anoxygenic purple phototrophic bacteria in Arctic soils. Soil Biology and Biochemistry 74: 193-200. doi: 10.1016/j.soilbio.2014.03.014

89. Feng Y, Lin X, Zhang J, Mao T, Zhu J (2011c) Soil purple phototrophic bacterial diversity under double cropping (rice-wheat) with free-air CO2 enrichment (FACE). European Journal of Soil Science 62: 533-540. doi: 10.1111/j.1365-2389.2011.01357.x

90. Waidner LA, Kirchman DL (2008) Diversity and Distribution of Ecotypes of the Aerobic Anoxygenic Phototrophy Gene pufM in the Delaware Estuary. Applied and Environmental Microbiology 74: 4012-4021. doi: 10.1128/aem.02324-07

91. Masin M, Nedoma J, Pechar L, Koblizek M (2008) Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. Environmental Microbiology 10: 1988-1996. doi: 10.1111/j.1462-2920.2008.01615.x

92. Peeters K, Ertz D, Willems A (2011) Culturable bacterial diversity at the Princess Elisabeth Station (Utsteinen, Sør Rondane Mountains, East Antarctica) harbours many new taxa. Systematic and Applied Microbiology 34: 360-367. doi: 10.1016/j.syapm.2011.02.002

93. Allgaier M, Uphoff H, Felske A, Wagner-Döbler I (2003) Aerobic Anoxygenic Photosynthesis in Roseobacter Clade Bacteria from Diverse Marine Habitats. Applied and Environmental Microbiology 69: 5051-5059. doi: 10.1128/aem.69.9.5051-5059.2003

94. Campbell BJ, Waidner LA, Cottrell MT, Kirchman DL (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. Environmental Microbiology 10: 99-109. doi: 10.1111/j.1462-2920.2007.01436.x

95. Wu L, Cui Y, Chen S (2011) Diversity of nifH gene sequences in the sediments of South China Sea. African Journal of Microbiology Research 5: 5972-5977. doi: 10.5897/ajmr11.1173

96. Nagashima K, Hiraishi A, Shimada K, Matsuura K (1997) Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. Journal of Molecular Evolution 45: 131-136. doi: 10.1007/pl00006212

97. Beja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser CM, DeLong EF (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature 415: 630-633.